

Taxonomies of networks from community structure

Jukka-Pekka Onnela,^{1,2,3,4,*} Daniel J. Fenn,^{5,4,*} Stephen Reid,³ Mason A. Porter,^{6,4} Peter J. Mucha,⁷ Mark D. Fricker,^{8,4} and Nick S. Jones^{3,9,4}

¹*Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA*

²*Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA*

³*Department of Physics, University of Oxford, Oxford OX1 3PU, UK*

⁴*CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK*

⁵*Mathematical and Computational Finance Group, University of Oxford, Oxford OX1 3LB, UK*

⁶*Oxford Centre for Industrial and Applied Mathematics,*

Mathematical Institute, University of Oxford, OX1 3LB, UK

⁷*Carolina Center for Interdisciplinary Applied Mathematics,*

Department of Mathematics and Institute for Advanced Materials,

Nanoscience & Technology, University of North Carolina, Chapel Hill, NC 27599, USA

⁸*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK*

⁹*Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK*

The study of networks has become a substantial interdisciplinary endeavor that encompasses myriad disciplines in the natural, social, and information sciences. Here we introduce a framework for constructing taxonomies of networks based on their structural similarities. These networks can arise from any of numerous sources: they can be empirical or synthetic, they can arise from multiple realizations of a single process (either empirical or synthetic), they can represent entirely different systems in different disciplines, etc. Because mesoscopic properties of networks are hypothesized to be important for network function, we base our comparisons on summaries of network community structures. Although we use a specific method for uncovering network communities, much of the introduced framework is independent of that choice. After introducing the framework, we apply it to construct a taxonomy for 746 networks and demonstrate that our approach usefully identifies similar networks. We also construct taxonomies within individual categories of networks, and we thereby expose nontrivial structure. For example, we create taxonomies for similarity networks constructed from both political voting data and financial data. We also construct network taxonomies to compare the social structures of 100 Facebook networks and the growth structures produced by different types of fungi.

Keywords: networks; clustering; community structure

I. INTRODUCTION

Although there is a long tradition of scholarship on networks, the last two decades have witnessed substantial advances in network science due to developments in physics, mathematics, computer science, sociology, and numerous other disciplines [1, 2]. Given that the questions asked by researchers in different fields can be surprisingly similar, it would be useful to be able to highlight similarities in network structures across disciplines in a systematic way. One way to approach this is to formulate a suitable means of comparing networks and to use this means to develop taxonomies of networks. Such taxonomies have the potential to facilitate the identification of problems from different disciplines that might be approached similarly in terms of both empirical analyses and theoretical modeling. For example, if a biological network depicting covariation of neural activity in different regions of the brain is demonstrated to be structurally similar to a financial network representing correlations of stock returns, then certain types of edge thresholding methods or structural null models might be applicable to

both situations.

From a historical perspective, classification of objects has often been central to the progress of science, as demonstrated by the periodic table of elements in chemistry and phylogenetic trees of organisms in biology [3]. It is plausible that an organization of networks has the potential to shed light on mechanisms for generating networks, reveal how an unknown network should be treated once one has discerned its position in a taxonomy, or help identify a network family's anomalous members. Further potential applications of network taxonomies include unsupervised study of multiple realizations of a given model process (e.g., characterizing the similarities and differences of many different networks drawn from the Erdős-Rényi random graph model using the same parameter values), examination of multiple empirical networks with known similar origins or generative processes, and the detection of anomalous changes in temporally ordered series of networks. In this paper, we develop a framework for the creation of network taxonomies [4]. In so doing, we develop the requisite diagnostic tools and discuss several case studies that suggest how our methodology can help illuminate relationships both between and within families of networks.

In aiming to construct taxonomies of networks, one

* These authors contributed equally to this work.

has to consider the scales at which one wants to compare differences in network structures. Much research has focused on extremes—either microscopic (e.g., node degree) or macroscopic (e.g., mean geodesic distance) properties—and numerous researchers have, for example, reported that many empirical networks possess heavy-tailed degree distributions or the small-world property [1, 5]. Given the ubiquity of such findings, it is clear that more nuanced approaches are needed to make useful comparisons between networks. Indeed, interpretations of microscopic and macroscopic approaches often implicitly assume that networks are homogeneous and ignore “mesoscopic” structures in networks. To overcome some of these limitations, earlier work has focused on the statistics of small, *a priori* specified modules called “motifs” [32, 37], role-to-role connectivity profiles of nodes [8], the isolation of statistically significant structures called “backbones” [9], interrelations of network modules [10], examination of the number of nodes located within “shells” [11], and the self-similarity of networks as characterized by fractal exponents [12]. The taxonomic framework that we develop in the present paper builds on the idea of examining network modules by computing community structures [13, 14], as was also done in the work of [15], and we subsequently compare signatures derived from community structure across networks. Importantly, although we use a specific method to uncover network communities, much of the introduced framework is independent of that choice. Consequently, our comparative framework can accommodate a large variety of community detection schemes.

The remainder of this paper is organized as follows. First, we discuss the detection of communities in networks in order to find coherent groups of nodes that are densely connected to each other. We then introduce *mesoscopic response functions* (MRFs), which allow us to probe how the community structure of a network changes as a function of a resolution parameter that determines network scales of interest. We then illustrate MRFs using several examples of networks and compare the MRFs for several well-known generative models of networks. We use MRFs to develop a means to measure distance between a pair of networks, and use this comparative measure to cluster networks and thereby develop taxonomies. Using 746 networks from numerous different fields, we construct a taxonomy of these networks. We then construct taxonomies of networks within fields using several case studies: voting in the United States Senate, voting in the United Nations General Assembly, Facebook networks at US universities, fungal networks, and networks of stock returns in the New York Stock Exchange. In each example, we expose structure that is either illuminating or can be checked against information from an external source (e.g., previously published investigations). This suggests that our method for comparing networks is capturing important similarities and differences. We conclude with a brief summary and discussion of our results. In addition, we provide further details

in the Appendices and Supplemental Material. Among other topics, we examine the robustness of the obtained taxonomies, address some computational issues, tabulate some of the basic properties of the networks that we investigated, and provide references for the network data sources used in this study.

II. MULTI-RESOLUTION COMMUNITY DETECTION

Our approach is based on network *community structure* [13, 14]. A community consists of a set of nodes for which there are more edges (or, in the case of weighted networks, a greater total edge weight) connecting the nodes in the set than what would be expected by chance. The algorithmic detection of communities is a particularly active area of network science, in part because communities are thought to be related to functional units in many networks and in part because they can strongly influence dynamical processes that operate on networks [13, 14].

In this paper, we detect communities using the multi-resolution Potts method [13, 14, 16], a generalization of modularity optimization [9, 13, 14, 16, 18, 19]. (Modularity optimization is perhaps the most popular approach for detecting communities.) Given a network adjacency matrix A_{ij} , we find communities by minimizing the Hamiltonian of the infinite-range N -state Potts spin glass

$$\begin{aligned}\mathcal{H}(\lambda) &= - \sum_{i \neq j} J_{ij}(\lambda) \delta(C_i, C_j) \\ &= - \sum_{i \neq j} (A_{ij} - \lambda P_{ij}) \delta(C_i, C_j),\end{aligned}\quad (1)$$

where C_i indicates the community (state) of node (spin) i , λ is a resolution parameter, and $\mathbf{J}(\lambda)$ is the coupling matrix with entries $J_{ij}(\lambda)$ representing the interaction strength between node i and node j in the Potts Hamiltonian. We use the (undirected-network) null model $P_{ij} = k_i k_j / (2m)$, where k_i denotes the strength (total edge weight) of node i and m is the total edge weight in the network [9]. By tuning the resolution parameter λ , we can detect communities at multiple scales of a network. Our particular choice of J_{ij} implies that we are optimizing modularity (with the addition of the resolution parameter) [13, 14].

To compare networks, we create profiles of summary statistics that characterize the community structure of each network at different mesoscopic scales. We also study a wide variety of networks that contain different numbers of nodes and edges. (We enumerate the networks that we consider in Table II of the Supplemental Material.) To ensure that we can compare the profiles for different networks, we sweep the resolution parameter λ from a minimum value Λ_{\min} to a maximum value Λ_{\max} (discussed in detail below). We define these quantities separately for each network such that the number of communities η into which the network is partitioned

is 1 at Λ_{\min} and is equal to the total number of nodes N at Λ_{\max} . In other words, one can think of λ as a parameter that controls the fragmentation of a network into communities.

To find the minimum and maximum resolution-parameter values, consider the interactions in Eq. (1). An interaction is called *ferromagnetic* when $J_{ij} > 0$ and *antiferromagnetic* when $J_{ij} < 0$. For each pair of nodes i and j , we find the resolution $\lambda = \Lambda_{ij}$ at which the interaction J_{ij} is neutral (i.e., $J_{ij}(\Lambda_{ij}) = 0$), leading to $\Lambda_{ij} = A_{ij}/P_{ij}$. We thereby identify two special resolutions:

$$\Lambda_{\min} = \max_{ij} \{\Lambda_{ij} | \eta(\lambda) = 1\}, \quad (2)$$

$$\Lambda_{\max} = \max_{ij} \{\Lambda_{ij}\} + \epsilon, \quad (3)$$

where $\epsilon > 0$ is any small number (we use $\epsilon = 10^{-6}$ in the present paper). The resolution Λ_{\min} is the largest Λ_{ij} value for which community detection yields a single community; note that this need not be the minimum non-zero value of Λ_{ij} . Including the small number ϵ in the definition of Λ_{\max} ensures that all edges are antiferromagnetic at resolution $\lambda = \Lambda_{\max}$ and thereby forces each node into its own community.

III. MESOSCOPIC RESPONSE FUNCTIONS (MRFS)

To describe how a network disintegrates into communities as the value of λ is increased from Λ_{\min} to Λ_{\max} (see Fig. 1(a) for a schematic), one needs to select summary statistics. There are many possible ways to summarize such a disintegration process, and we focus on three diagnostics that characterize fundamental properties of network communities.

First, we use the value of the Hamiltonian $\mathcal{H}(\lambda)$ (1), which is a scalar quantity closely related to network modularity and quantifies the energy of the system [13, 14]. Second, we calculate a partition entropy $S(\lambda)$ to characterize the community size distribution. To do this, let n_k denote the number of nodes in community k and define $p_k = n_k/N$ to be the probability to choose uniformly at random a member node of community k . This yields a (Shannon) partition entropy of $S(\lambda) = -\sum_{k=1}^{\eta(\lambda)} p_k \log p_k$, which quantifies the disorder in the associated community size distribution. Third, we use the number of communities $\eta(\lambda)$.

Needing to normalize \mathcal{H} , S , and η to compare them effectively across networks, we define an *effective energy*

$$\mathcal{H}_{\text{eff}}(\lambda) = \frac{\mathcal{H}(\lambda) - \mathcal{H}_{\min}}{\mathcal{H}_{\max} - \mathcal{H}_{\min}} = 1 - \frac{\mathcal{H}(\lambda)}{\mathcal{H}_{\min}}, \quad (4)$$

where $\mathcal{H}_{\min} = \mathcal{H}(\Lambda_{\min})$ and $\mathcal{H}_{\max} = \mathcal{H}(\Lambda_{\max})$; an *effective entropy*

$$S_{\text{eff}}(\lambda) = \frac{S(\lambda) - S_{\min}}{S_{\max} - S_{\min}} = \frac{S(\lambda)}{\log N}, \quad (5)$$

where $S_{\min} = S(\Lambda_{\min})$ and $S_{\max} = S(\Lambda_{\max})$; and an *effective number of communities*

$$\eta_{\text{eff}}(\lambda) = \frac{\eta(\lambda) - \eta_{\min}}{\eta_{\max} - \eta_{\min}} = \frac{\eta(\lambda) - 1}{N - 1}, \quad (6)$$

where $\eta_{\min} = \eta(\Lambda_{\min})$ and $\eta_{\max} = \eta(\Lambda_{\max})$.

Some networks contain a small number of entries Λ_{ij} that are orders-of-magnitude larger than most other entries. For example, in the network of Facebook friendships at Caltech [21], 98% of the Λ_{ij} entries are less than 100, but 0.02% of them are larger than 8000. These large Λ_{ij} values arise when two low-strength nodes become connected. Using the null model $P_{ij} = k_i k_j / (2m)$, the interaction between two nodes i and j becomes antiferromagnetic when $\lambda > A_{ij}/P_{ij} = 2m A_{ij} / (k_i k_j)$. If the network has a large total edge weight but both i and j have small strengths compared to other nodes in the network, then λ needs to be large to make the interaction antiferromagnetic. In prior studies, network community structure has been investigated at different mesoscopic scale by considering plots of various diagnostics as a function of the resolution parameter [13, 14, 16]. In the present example, such plots would be dominated by interactions that require large resolution-parameter values to become antiferromagnetic. To overcome this issue, we define the *effective fraction of antiferromagnetic edges*

$$\xi = \xi(\lambda) = \frac{\ell^A(\lambda) - \ell^A(\Lambda_{\min})}{\ell^A(\Lambda_{\max}) - \ell^A(\Lambda_{\min})} \in [0, 1], \quad (7)$$

where $\ell^A(\lambda)$ is the total number of antiferromagnetic interactions for the given value of λ in the network. In other words, it is the number of Λ_{ij} elements that are smaller than λ . Thus, $\ell^A(\Lambda_{\min})$ is the largest number of antiferromagnetic interactions for which the network still forms a single community, and the effective number of antiferromagnetic interactions $\xi(\lambda)$ is the number of antiferromagnetic interactions (normalized to the unit interval) in excess of $\ell^A(\Lambda_{\min})$. The function $\xi(\lambda)$ increases monotonically in λ .

Sweeping λ from Λ_{\min} to Λ_{\max} corresponds to sweeping the value of ξ from 0 to 1. (One can think of λ as a continuous variable and ξ as a discrete variable that changes with events.) As we perform such sweeping for a given network, the number of communities increases from $\eta(\xi = 0) = 1$ to $\eta(\xi = 1) = N$ and yields a vector $(\mathcal{H}_{\text{eff}}(\xi), S_{\text{eff}}(\xi), \eta_{\text{eff}}(\xi))$ whose components we call the *mesoscopic response functions* (MRF) of that network. Because $\mathcal{H}_{\text{eff}} \in [0, 1]$, $S_{\text{eff}} \in [0, 1]$, $\eta_{\text{eff}} \in [0, 1]$, and $\xi \in [0, 1]$ for every network, we can compare the MRFs across networks and use them to identify groups of networks with similar mesoscopic structures. In Fig. 1(b), we show the Zachary Karate Club network [83] for different values of ξ . As more edges become antiferromagnetic, the network fragments into smaller communities, and panel (c) shows the corresponding MRFs. In Fig. 2, we show a schematic of the MRF in which we emphasize its interpretation as a 3-dimensional vector. In Fig. 3, we show example MRFs for several other networks.

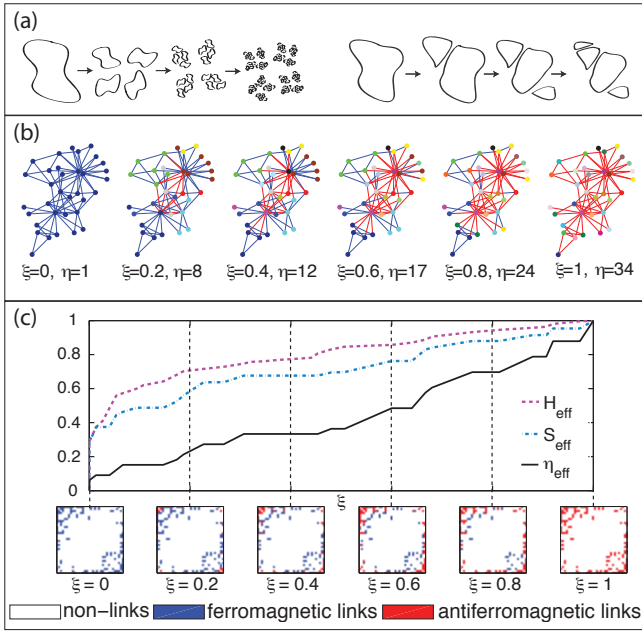


FIG. 1. (Color online) (a) Schematic of some of the ways that a network can break up into communities as the value of λ (or ξ) is increased. (b) Zachary Karate Club network [83] for different values of the effective fraction of antiferromagnetic edges ξ . All interactions are either ferromagnetic or antiferromagnetic, i.e. for the values of ξ used, there are no neutral interactions. We color edges in blue if the corresponding interactions are ferromagnetic, and we color them red if the interactions are antiferromagnetic. We color the nodes based on community affiliation. (c) The \mathcal{H}_{eff} , \mathcal{S}_{eff} , and η_{eff} MRFs, and the interaction matrix \mathbf{J} for different values of ξ . We color elements of the interaction matrix by depicting the absence of an edge in white, ferromagnetic edges in blue, and antiferromagnetic edges in red.

Although minimizing Eq. (1) is an NP-hard problem [23] and \mathcal{H} possesses a complicated landscape of local optima for many networks [24], there exist numerous good computational heuristics that make finding a nearly-optimal partition of the network into communities at a given resolution computationally tractable [13, 14]. Thus far, we have reported results that were obtained by optimizing modularity using the locally greedy Louvain algorithm [25] because its speed was important for studying large networks. We have compared the results that we report in the present work to those obtained from optimizing modularity using spectral and simulated-annealing algorithms, and obtained similar MRFs and taxonomies for them (see Appendix B2 for more details).

IV. EXAMPLES OF MRFs

The shapes of the MRFs summarize many factors—including the fraction of possible edges in a network that are actually present, the relative weights of inter- versus

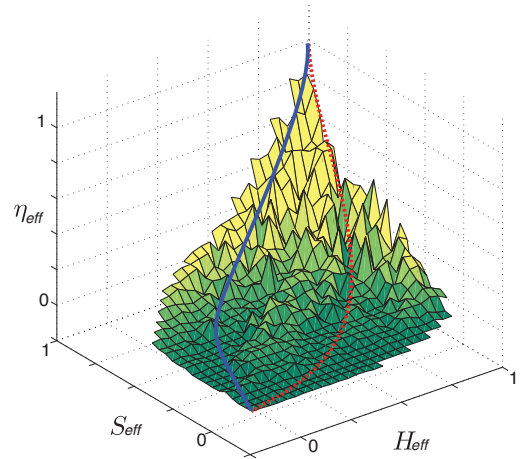


FIG. 2. (Color online) The mesoscopic response function (MRF) of a given network consists of a 3-dimensional vector $(\mathcal{H}_{\text{eff}}(\xi), \mathcal{S}_{\text{eff}}(\xi), \eta_{\text{eff}}(\xi))$, where $\xi \in [0, 1]$. By construction, the MRF starts from the bottom front corner $[\mathcal{H}_{\text{eff}}(\xi = 0), \mathcal{S}_{\text{eff}}(\xi = 0), \eta_{\text{eff}}(\xi = 0)]$ and ends at the top back corner $[\mathcal{H}_{\text{eff}}(\xi = 1), \mathcal{S}_{\text{eff}}(\xi = 1), \eta_{\text{eff}}(\xi = 1)]$. The colored surface plot shows where most MRFs lie. We also show schematic MRFs in blue (solid curve) and red (dashed curve).

intra-community edges, the edge weights compared with the expected edge weights in the null model, the number of edges that need to become antiferromagnetic for a community to fragment, and the way in which the communities fragment (e.g., whether a community splits in half or a single node leaves a community when a particular edge becomes antiferromagnetic). To understand the effects of some of these factors on the shapes of the MRFs, we consider some examples.

Of particular interest are plateaus in the η_{eff} and \mathcal{S}_{eff} curves that are accompanied by large increases in \mathcal{H}_{eff} . As illustrated in panel 3(a), the New York Stock Exchange (NYSE) network from 1980 to 1999 [22] provides a good example of this behavior. This network is an instance from the category of *similarity networks*. We use this label to describe networks that have been constructed by starting from some node-level quantity or attribute and then defining the edges based on some form of similarity or correlation measure between each pair of nodes. Similarity networks tend to be complete (or almost complete) and weighted networks, except when they have been deliberately thresholded. In this particular example, each node represents a stock, and the strength of the edge connecting stocks i and j is linear in the Pearson correlation between the daily logarithmic returns of the stocks. (See Section IX E for more details.) Plateaus imply that as the resolution λ is increased (leading to an increase in \mathcal{H}_{eff}), the communities remain unchanged even though the number and strength of antiferromagnetic interactions increase. As λ is increased and more interactions become antiferromagnetic, there is an

increased energy incentive for communities to break up. Community partitions in such plateaus tend to be robust and have the potential to represent interesting structures [13, 14, 16, 27].

In Fig. 3(b), we show MRFs for a “fractal” network [4], which demonstrates that plateaus in the η_{eff} and S_{eff} curves need not be accompanied by significant changes in \mathcal{H}_{eff} . Such plateaus can be explained by considering the distribution of Λ_{ij} values. If several interactions have identical values of Λ_{ij} , then the interactions all become antiferromagnetic at exactly the same resolution value. This leads to a significant increase in the effective fraction of antiferromagnetic edges ξ but only a small change in \mathcal{H}_{eff} . If these interactions do not result in additional communities, then we obtain plateaus in the η_{eff} and S_{eff} curves.

To demonstrate qualitatively different behavior, we show the MRFs for the Biogrid *Drosophila melanogaster* network and the Garfield Scientometrics citation network in Fig. 3(c) and Fig. 3(d), respectively. A common feature in these MRFs is the sharp initial increase in the curves that results from the networks initially breaking into two communities.

Another family of networks, which we will discuss in more detail in our case studies, are political voting networks. These voting networks are also similarity networks: we have constructed these networks so that an edge between two nodes indicates the level of agreement on votes between two entities, and each edge takes a value between 0 and 1. In Fig. 3(e), we show the MRFs for the voting network of the United Kingdom House of Commons during the period 2001–2005 [53]; in Fig. 3(f), we show the MRFs for the roll-call voting network for the 108th (2003–2004) United States House of Representatives [30, 50–52]. In both cases, we observe that sharp increases in \mathcal{H}_{eff} can be accompanied by only small changes in η_{eff} and S_{eff} . To see how this can arise, we again consider the distribution of Λ_{ij} values. If the Λ_{ij} distribution is multi-modal, there can be a large difference between consecutive Λ_{ij} values. A large increase in λ is then needed to increase ξ , which in turn results in a large change in \mathcal{H}_{eff} . However, the change in η_{eff} is small because this only results in a single additional antiferromagnetic interaction.

V. COMPARING NETWORK MODELS

To provide further insights into MRFs, we consider Erdős-Rényi (ER) [1], Barabási-Albert (BA) [3], and Watts-Strogatz (WS) [2] networks. These network models are stochastic, and there is a large ensemble of possible network realizations for each choice of parameter values in these models. However, even with the ensuing structural variation, networks generated by a given one of these three models exhibit similar properties at mesoscopic and macroscopic scales, so we expect MRFs for different realizations of a given model to be similar.

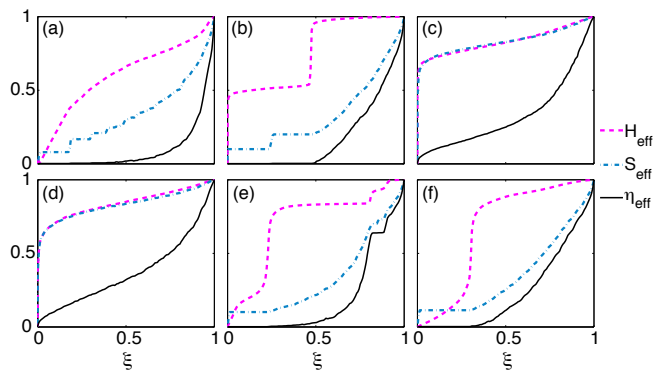


FIG. 3. (Color online) Example mesoscopic response functions (MRFs). The curves show \mathcal{H}_{eff} (pink, dashed), S_{eff} (blue, dash-dotted), and η_{eff} (black, solid) as a function of the effective fraction of antiferromagnetic edges ξ for the following networks: (a) New York Stock Exchange (NYSE), 1980–1999 [22]; (b) Fractal (10,2,8) [4]; (c) Biogrid *D. melanogaster* [55]; (d) Garfield scientometrics citations [40]; (e) United Kingdom House of Commons voting, 2001–2005 [53]; (f) Roll-call voting of 108th United States House of Representatives [30, 50–52].

In Fig. 4, we compare the MRFs for 1000 realizations of each model for networks with $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. For the WS networks, we set the edge rewiring probability at $p = 0.1$. As illustrated in Fig. 4, we obtain a narrow range of possible MRFs for fixed parameter values. This comparison illustrates that the MRF profiles of the three different models are distinctive. In addition, for each model there is little variation in the behavior of the MRFs across different network realizations with the same parameter values.

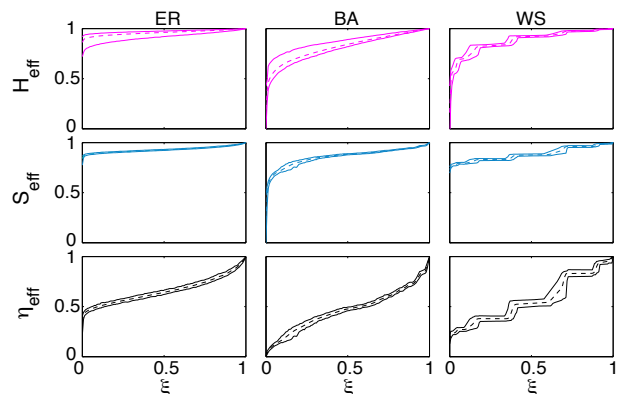


FIG. 4. (Color online) MRFs for 1000 realizations of Erdős-Rényi (ER), Barabási-Albert (BA), and Watts-Strogatz (WS) networks. Each network has $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. For each value of ξ , the upper curves show the maximum values of \mathcal{H}_{eff} (top row), S_{eff} (middle row), and η_{eff} (bottom row) for all networks in the ensemble; the lower curves show the corresponding minimum value, and the dashed curves show the corresponding mean.

It is also instructive to consider variation in MRF shapes for a particular network model for different parameter values. We focus on WS networks because they illuminate the effect of the distribution of Λ_{ij} values on the shapes of the MRFs. In Fig. 5, we show MRFs for WS networks for different values of the edge rewiring probability p . (We continue using $N = 1000$ and $\langle k \rangle = 10$.) We also show the distribution of Λ_{ij} values for each network.

For small rewiring probabilities, the MRFs have lots of steps. As with prior examples, we can see how this feature arises by considering the distribution of Λ_{ij} values. When the rewiring probability is small, many nodes possess the same degree, which results in the presence of many interactions with identical Λ_{ij} values (see the bottom left panel of Fig. 5). Because several interactions have identical Λ_{ij} values, these interactions all become antiferromagnetic at exactly the same resolution-parameter value, so the behavior of MRFs only changes for a small number of ξ values. As the rewiring probability p is increased, the degree and Λ_{ij} distributions become more heterogeneous, which leads to smoother MRFs. For a rewiring probability of $p = 1$, the WS network is just an ER network.

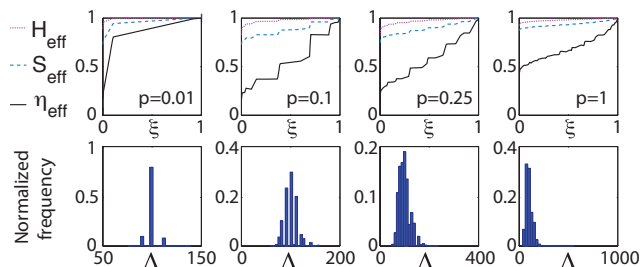


FIG. 5. (Color online) Upper panels: MRFs for Watts-Strogatz networks for different values of the rewiring probability p . Each network has $N = 1000$ nodes and mean degree $\langle k \rangle = 10$. Lower panels: distributions of Λ_{ij} values for each network. As expected, the MRFs for $p = 1$ are identical to those of an Erdős-Rényi network with $N = 1000$ and $\langle k \rangle = 10$.

VI. MEASURING DISTANCE BETWEEN NETWORKS

In the framework that we have introduced in this paper, comparing two networks at the mesoscopic level amounts to characterizing the differences in behavior of the corresponding MRFs. To quantify such differences, we define a distance between two networks with respect to one of the summary statistics as the area between the corresponding MRFs. For example, the distance between two networks i and j with respect to the effective energy \mathcal{H}_{eff} is given by

$$d_{ij}^{\mathcal{H}} = \int_0^1 |\mathcal{H}_{\text{eff}}^i(\xi) - \mathcal{H}_{\text{eff}}^j(\xi)| d\xi. \quad (8)$$

For the effective entropy and effective number of communities, the distances are given by $d_{ij}^S = \int_0^1 |S_{\text{eff}}^i(\xi) - S_{\text{eff}}^j(\xi)| d\xi$ and $d_{ij}^\eta = \int_0^1 |\eta_{\text{eff}}^i(\xi) - \eta_{\text{eff}}^j(\xi)| d\xi$, respectively.

We represent the resulting three sets of distances (computed for each pair of networks from the 746 networks that we consider, see Table I) in matrix form as $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^η . These distance measures have several desirable properties. First, they compare MRFs across all network scales (i.e., for all values of ξ); second, each distance is bounded between 0 and 1; third, the distances are easy to interpret, as each of them corresponds to the geometric area between (a certain dimension of) a pair of MRFs; and finally, we find *a posteriori* that these distances can be used to cluster networks accurately (see the discussions below).

We have computed MRFs for the energy \mathcal{H} , entropy S , and number of communities η , but we can proceed similarly with any desired summary statistic. If two diagnostics provide similar information, then one of them can be excluded without significant loss of information. We checked whether the summary statistics were sufficiently different, for the set of networks considered here, for it to be worthwhile to include all of them by calculating the Pearson correlation coefficient between their corresponding distance measures. The correlations between the pairs of distances are $r(d_{ij}^{\mathcal{H}}, d_{ij}^S) \doteq 0.36$, $r(d_{ij}^{\mathcal{H}}, d_{ij}^\eta) \doteq 0.24$, and $r(d_{ij}^S, d_{ij}^\eta) \doteq 0.58$. These correlations are not sufficiently high to justify excluding any of the summary statistics.

In the interest of parsimony—and given the non-vanishing correlations between the distance measures—we reduce the number of distance measures using principal component analysis (PCA) [39]. Starting with \mathcal{N} networks, we create a $\frac{1}{2}\mathcal{N}(\mathcal{N} - 1) \times 3$ matrix in which each column corresponds to the vector representation of the upper triangle of one of the distance matrices $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , \mathbf{D}^η , and we perform a PCA on this matrix. We then define a distance matrix \mathbf{D}^p with elements $d_{ij}^p = w_{\mathcal{H}}d_{ij}^{\mathcal{H}} + w_S d_{ij}^S + w_\eta d_{ij}^\eta$, where the weights are the coefficients for the first principal component, and we normalize the sum of squared coefficients to unity. The coefficients are $w_{\mathcal{H}} \doteq 0.24$, $w_S \doteq 0.79$, and $w_\eta \doteq 0.57$. The first component accounts for about 69% of the variance, so the distances \mathbf{D}^p provide a reasonable single-variable projection of the distances $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^η .

It is important that the distance measures for comparing networks are robust to small perturbations in network structure. Because many of the networks that we study are constructed empirically, they might contain false positives and false negatives. In other words, the networks might falsely identify a relationship where none exists, and they also might fail to identify an existing relationship. Consequently, the topology and edge weights of an observed network might be slightly different than those of the actual underlying network. To test the robustness of our distance measures to such observational errors, we recalculate the MRFs for a subset of relatively small un-

weighted networks in which, for each network, we rewire a number of edges corresponding to a given percentage of the total number of edges (5%, 10%, 20%, 50%, or 100%). See Appendix A for more details. (We study networks with up to 1000 nodes and only consider a subset of 25 networks because of the computational costs of rewiring a large number of networks multiple times; however, we have performed the same investigation for 5 different subsets of 25 networks and obtained similar results. We list the networks in each subset in Table I of the Supplemental Material.) We investigate two rewiring mechanisms: one in which the degree distribution is maintained, where we also ensure after each rewiring that the network forms a single connected component; and another in which the only constraint is that the network continues to consist of a single connected component after each edge rewiring [40]. We find in both cases that the structures of the block-diagonalized distance matrices for the 25 networks (see Figs. 14 and 15 in Appendix A) are robust to random perturbations of the networks, thereby suggesting that our MRF distance measures are not sensitive to small structural perturbations.

VII. CLUSTERING NETWORKS

We assign each of the 746 networks to a category based on its type (see Table I). Due to the varying availability of different types of network data, the included networks are not evenly distributed across these categories. Many of the networks are either different temporal snapshots of the same system or different realizations of the same type of network. To have a more balanced distribution across the different categories, we focus on 189 of the 746 networks. We only include categories for which we have 8 or more networks, and we selected a subset of networks (uniformly at random) from the larger categories. We also exclude all synthetic networks. See Section IV of the Supplemental Material for the list of networks that we consider and Fig. 1 in Section II of the Supplemental Material for a dendrogram showing a taxonomy we constructed using all 746 networks.

Our primary reason for assigning each network to a category is to use such an external categorization to help assess the quality of taxonomies produced by the unsupervised MRF clustering. For each way of computing distance, we construct a dendrogram for the set of networks using average linkage clustering, which is an agglomerative hierarchical clustering technique [13, 41][42]. In Fig. 6, we show a dendrogram obtained from the distance matrix \mathbf{D}^p . The colored rectangle underneath each leaf indicates the network category. Contiguous blocks of color demonstrate that networks from the same category have been grouped together using the MRF clustering method, and the presence of such contiguous color blocks is an indication of the success of the MRF clustering scheme.

The assignment of the networks to one of these cate-

TABLE I. Network categories, the total number of networks assigned to each category, and the number of networks from each category included in the taxonomy in Fig. 6. For the full taxonomy that uses all 746 networks, see Fig. 1 of the Supplemental Material.

Category	All networks	Taxonomy networks
Political: voting	285	23
Facebook	100	15
Fungal	65	12
Synthetic	58	0
Financial	54	6
Metabolic	43	15
Social	26	26
Political: cosponsorship	26	26
Other	23	0
Protein interaction	22	22
Political: committee	16	16
Brain	12	12
Language	8	8
Collaboration	8	8
Total	746	189

gories is of course to some extent subjective, as several of the networks could belong to more than one category. For example, we could categorize the network of jazz musicians [20] as either a collaboration network or a social network. The initial selection of network categories is also somewhat subjective. One could argue that if one has a social network category, then it is not necessary to have a collaboration network category as well because a collaboration network is a type of social network. We have attempted to maintain a balance between having too many categories and having too few of them. When such ambiguities have arisen, we have systematically chosen the more specific of the relevant categories (e.g., we placed the jazz musician network in the category of collaboration networks rather than in the category of social networks).

VIII. TAXONOMIES OF EMPIRICAL NETWORKS

All of the networks in some categories appear in blocks of adjacent leaves in the dendrogram in Fig. 6. For example, there is a cluster of political voting networks at the far left of the dendrogram. This cluster includes voting networks from the US Senate, the US House of Representatives, the UK House of Commons, and the United Nations General Assembly (UNGA). The clustering of these voting networks suggests that there are some common features in the network representations of the different legislative bodies. We also obtain blocks that consist

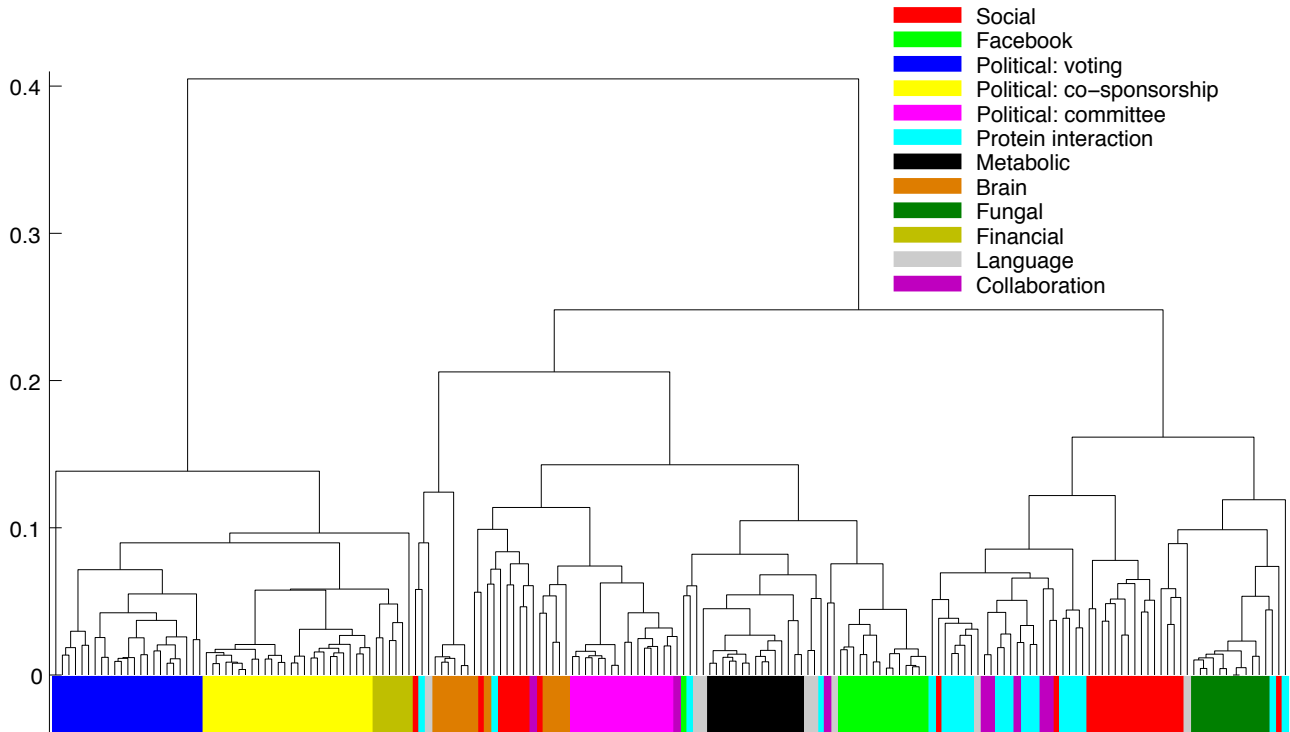


FIG. 6. (Color online) Taxonomy for 189 networks. We construct the dendrogram (tree) using the distance \mathbf{D}^p and average linkage clustering. We order the leaves of the dendrogram to minimize the distance between adjacent nodes and color the leaves to indicate the type of network.

of all political committee networks and all metabolic networks.

There are also several categories for which all except one or two networks cluster into a contiguous block. For example, all but two of the fungal networks appear in the same block and all but one of the Facebook networks are clustered together. The isolated Facebook network is the Caltech network, which is the smallest network of this type and which appears in a group next to that containing all of the other Facebook networks. We remark that the social organization of the community structure of the Caltech Facebook network has been shown to be different from those of the other Facebook networks [21].

Networks of certain categories do not appear in near-contiguous blocks. For example, protein interaction networks appear in several clusters. These networks represent interactions within several different organisms, so we would not expect all of them to be clustered together. Moreover, the data that we employed includes examples of protein interaction networks for the same organism in which the interactions were identified using different experimental techniques, and these networks do not cluster together. This supports previous work suggesting that the properties of protein interaction networks are very sensitive to the experimental procedure used to identify

the interactions [44, 45]. Social networks are also distributed throughout the dendrogram. This is unsurprising given the extremely broad nature of the category, which includes networks of very different sizes with edges representing a diverse range of social interactions. The leftmost outlying social network is the network of Marvel comic book characters [72], which is arguably an atypical social network.

The grouping (and, to some extent, the non-grouping) of networks by category suggests that the PCA-distance \mathbf{D}^p between MRFs of different networks produces a sensible taxonomy. It is important to ask, however, whether a simpler approach based on a single network diagnostic, such as edge density, can be comparably successful at constructing a taxonomy. In Appendix D, we demonstrate using some well-known diagnostics that this does not appear to be the case, as the diagnostics we tried were unable to reproduce or explain the classifications that we produced using the MRFs.

In order to compare the aggregate shapes of the MRFs across categories, we show the bounds of the \mathcal{H}_{eff} , S_{eff} , and η_{eff} curves for each category in Fig. 7. We again consider all empirical network categories with at least 8 networks in them. This illustrates that the MRFs for some classes of networks (such as political cosponsor-

ship and metabolic networks) are very similar to each other, whereas there are large variations in the MRFs for other categories (such as social and protein interaction networks). The variety of different MRFs for the social and protein interactions is consistent with the fact that their constituent networks are scattered throughout the dendrogram in Fig. 6.

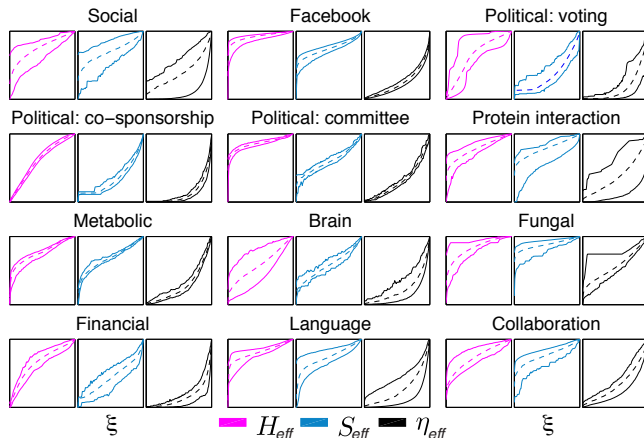


FIG. 7. (Color online) MRFs for all of the network categories containing at least 8 networks (see Table I). At each value of ξ , the upper curve shows the maximum value of \mathcal{H}_{eff} (pink, left panel in each category), \mathcal{S}_{eff} (blue, center panel), and η_{eff} (black, right panel) for all networks in the category and the lower curve shows the minimum value. The dashed curves show the corresponding mean MRFs.

IX. CASE STUDIES

We now consider several case studies, in which we generate taxonomies for multiple realizations of particular types of networks and multiple time slices of particular networks. This enables us to compare these networks and (in some cases) illustrate possible connections between network function and mesoscopic network structure.

A. Voting in the United States Senate

Our first example deals with roll-call voting in the United States Senate [30, 47, 50–52]. Establishing a taxonomy of networks detailing the voting similarities of individual legislators complements previous studies of these data, and it facilitates the comparison of voting similarity networks across time. We consider Congresses 1–110, which cover the period 1789–2008. As in Ref. [50], we construct networks from the roll-call data [30, 51] for each two-year Congress such that the adjacency matrix element $A_{ij} \in [0, 1]$ represents the number of times Senators i and j voted the same way on a bill (either both in

favor of it or both against it) divided by the total number of bills on which both of them voted. Following the approach of Ref. [51], we only consider “non-unanimous” roll call votes, which are defined as votes in which at least 3% of the Senators were in the minority.

Much research on the US Congress has been devoted to the ebb and flow of partisan polarization over time and the influence of parties on roll-call voting [50, 52]. In highly polarized legislatures, representatives tend to vote along party lines, so there are strong similarities in the voting patterns of members of the same party and strong differences between members of different parties. In contrast, during periods of low polarization, the party lines become blurred. The notion of partisan polarization can be used to help understand the taxonomy of Senates in Fig. 8, in which we consider two measures of polarization. The first measure uses DW-Nominate scores (a multi-dimensional scaling technique commonly used in political science [51, 52]), where the extent of polarization is given by the absolute value of the difference between the mean first dimension DW-Nominate scores for members of one party and the same mean for members of the other party [30, 51, 52]. In particular, we use the simplest such measure of polarization, called MPR polarization, which assumes a competitive two-party system and hence cannot be calculated prior to the 46th Senate. The second measure we consider is network modularity Q , which was recently shown to be a good measure of polarization even for Congresses without clear party divisions [50]. Modularity is given in terms of the energy \mathcal{H} in Eq. (1) by $Q = -\mathcal{H}(\lambda = 1)/(2m)$. These two measures exhibit fairly close agreement on the level of polarization of each Congress for which they can both be calculated [50].

In Fig. 8(a), we include bars under the dendrograms to represent the two polarization measures, both of which have been normalized to lie in the interval $[0, 1]$. The bars demonstrate that Senates with similar levels of polarization (measured in terms of both DW-Nominate scores and modularity values) are usually assigned to the same group, suggesting that our MRF clustering technique groups Senates based on the polarization of roll-call votes. We have also colored dendrogram groups according to their mean levels of polarization using modularity, where the brown group in the dendrogram corresponds to the most highly polarized Senates and the blue group corresponds to the least polarized Senates. Although one ought to expect similarity in the results from the modularity-based measure of polarization and the MRF clustering, it is important to stress that the MRF clustering method is based on different principles; modularity quantifies the extent to which a given network is “modular”, whereas the MRF clustering explicitly compares the differences in modular structures between any two networks at all scales.

In Fig. 8(a), we also show the clusters that we obtained for the Senate. They closely match the different periods of polarization that have been identified using

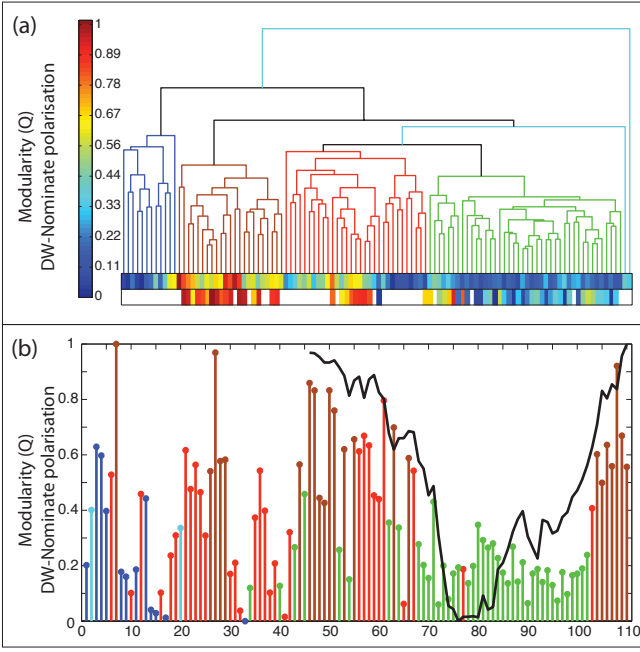


FIG. 8. (Color online) **(a)** Dendrogram for Senate roll-call voting networks for the 1st–110th Congresses. Each leaf in the dendrogram represents a single Senate. Two horizontal color bars below the dendrograms indicate polarization measured in terms of modularity (upper bar) and DW-Nominate scores (lower bar). We color the branches in the dendrogram corresponding to periods of similar polarization. **(b)** Polarization of the US Senate as a function of time. The height of each stem indicates the level of polarization measured using modularity, and the color of each stem gives the cluster membership of each Senate in (a). The black curve shows the DW-Nominate polarization. Note that we have rescaled both measures to the interval [0, 1].

modularity and DW-Nominate [50]. The cluster with the highest mean polarization (shown in brown) consists of Senates 7, 26–29, 44, 46–51, 53, 55, 66, and 104–110. The 104th–110th Congresses correspond to a period of extremely high polarization following the 1994 “Republican Revolution”, in which the Republican party earned majority status in the House of Representatives for the first time in more than 40 years [30, 50, 52]. The cluster with the second highest mean polarization (shown in red) includes several contiguous blocks of Senates, such as those from Congresses 21–25, 35–39, and 56–61. The 21st–25th Congresses (1829–1839) corresponded to a period of partisan conflict between supporters of John Quincy Adams and Andrew Jackson; it lasted until the emergence of the Whigs and the Democratic party in the 25th Congress [48, 50]. The American Civil War started during the 37th Congress, and a third party known as the Populist Party was strong during the 56th–58th Congresses.

The main differences between different clusters occur in the \mathcal{H}_{eff} response functions. For the most polarized Senates, there is a sharp shoulder in the \mathcal{H}_{eff} MRF that

becomes less pronounced as the polarization decreases. We illustrate this in Fig. 9, in which we compare the \mathcal{H}_{eff} MRFs for the (low-polarization) 85th and (high-polarization) 108th Senates. The shoulder in the \mathcal{H}_{eff} curve for the 108th Senate is very pronounced, which can be explained by considering the distribution of Λ_{ij} values. The 108th Senate has a bimodal Λ_{ij} distribution that contains a trough at $\Lambda_{ij} = 1$. Recall that $\Lambda_{ij} = A_{ij}/P_{ij}$, so Λ_{ij} compares the observed voting similarity A_{ij} of legislators i and j with the similarity $P_{ij} = k_i k_j / (2m)$ expected from random voting. If $\Lambda_{ij} < 1$, legislators i and j vote differently more frequently than expected (with respect to the chosen null model); if $\Lambda_{ij} > 1$, they vote more similarly than expected. Therefore, the peaks in the Λ_{ij} distribution above and below 1 correspond, respectively, to intra-party and inter-party voting blocs. In a Senate with low polarization, legislators from different parties often vote in the same manner, so the values of Λ_{ij} no longer separate two distinct types of behavior.

We also examined roll-call voting networks in the US House of Representatives and found many similar features as the ones that we have presented for the US Senate. For example, the highly polarized 104th–110th Congresses, which followed the “Republican Revolution”, appear in the same cluster for both the House and Senate. We also observed some differences in the clusters for the two chambers. For example, the 78th–102nd Senates all appeared in the same cluster. For the House, however, Congresses 80, 88, 89, and 98–102 did not appear in the same cluster as the other Congresses between 78 and 102; instead, they appeared in a cluster that also included the 26th–28th Houses. This was a particularly eventful period: the 25th Congress saw the emergence of the Whigs and the Democratic Party, and the abolitionist movement was also prevalent (e.g., the Amistad seizure occurred in 1839 during the 26th Congress).

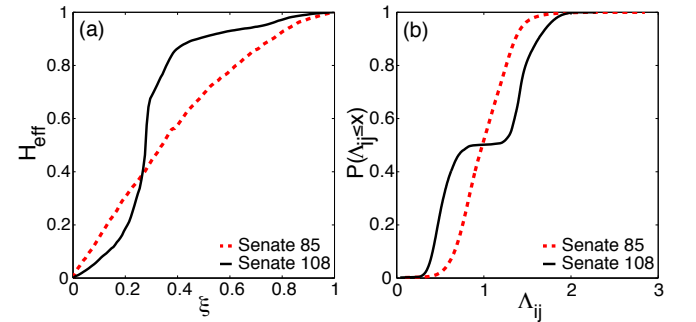


FIG. 9. (Color online) Comparison of the (low-polarization) 85th Senate and the (high-polarization) 108th Senate. The panels show **(a)** the \mathcal{H}_{eff} MRFs and **(b)** the cumulative distributions of Λ_{ij} values.

B. Voting in the United Nations General Assembly

The United Nations General Assembly (UNGA) is one of the principal organs of the United Nations (UN), and it is the only part of the UN in which all member nations have equal representation. Although most resolutions are neither legally nor practically enforceable because the General Assembly lacks enforcement powers on most issues, it is the only forum in which a large number of states meet and vote regularly on international issues. It also provides an interesting point of comparison with roll-call voting in the US Congress, as the level of agreement on UN resolutions tends to be much higher than that in the Senate and House [49].

We study voting for the 1st–63rd sessions (covering the period 1946–2008), where each session corresponds to a year [50]. For each session, we define an adjacency matrix \mathbf{A} whose elements A_{ij} represent the number of times countries i and j voted in the same manner in a session (i.e., the sum of the number of times both countries voted yea on the same resolution, both countries voted nay on the same resolution, or both countries abstained from voting on the same resolution) divided by the total number of resolutions on which the UNGA voted in a session. The matrix \mathbf{A} , with elements $A_{ij} \in [0, 1]$, thereby represents a (similarity) network of weighted edges between countries.

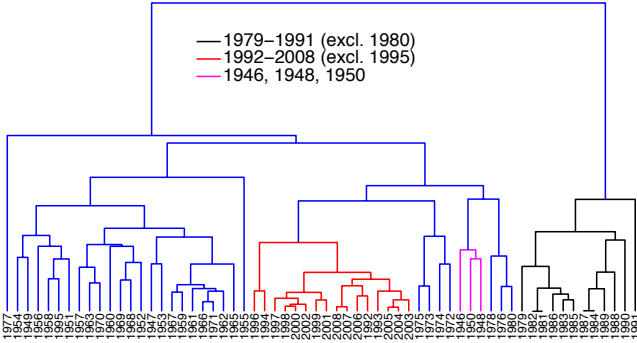


FIG. 10. (Color online) Dendrogram for the United Nations General Assembly resolution voting network for the 1st–63rd sessions (excluding the 19th session), covering the period 1946–2008. Each leaf in the dendrogram represents a single session. In the text, we discuss the coloring of groups of branches in the dendrogram.

We cluster UNGA sessions by comparing MRFs for the corresponding voting networks. In Fig. 10, we plot a dendrogram of the UNGA sessions and highlight some of the clusters, which correspond to notable periods in the recent history of international relations. The red cluster in the middle of the dendrogram consists of all post-Cold War sessions (1992–2008) except 1995. This group forms a larger cluster with some UNGA sessions from the 1970s and a cluster consisting of 1946, 1948, and 1950. These three sessions (shown in magenta) are all noteworthy:

1946 was the first session of the UNGA, the Universal Declaration of Human Rights was introduced during the 1948 session, and the “Uniting for Peace” resolution was passed during the 1950 session. At the rightmost part of the dendrogram, we color in black a group that consists of all sessions from 1979 to 1991 (excluding 1980). The beginning of this period marked the end of Détente between the Soviet Union and the US following the former’s invasion of Afghanistan at the end of 1979, and the end of this period saw the end of the Cold War. The large blue cluster in the leftmost part of the dendrogram consists primarily of sessions from before 1971 (though it also includes the sessions in 1977 and 1995).

C. Facebook

We now consider Facebook networks for 100 US universities [21]. The nodes in each network represent users of the Facebook social networking site, and the unweighted edges represent reciprocated “friendships” between users at a single-time snapshot in September 2005. We consider only edges between students at the same university, as this allows us to compare the structure of the networks at the different institutions. These networks represent complete data sets obtained directly from Facebook. In contrast to the previous examples, we are not comparing snapshots of the same network at different times but are instead comparing multiple realizations of the same type of network that have evolved independently. Such real-world ensembles of network data are rare, and constructing a taxonomy will hopefully allow us to compare and contrast the social organization at these institutions.

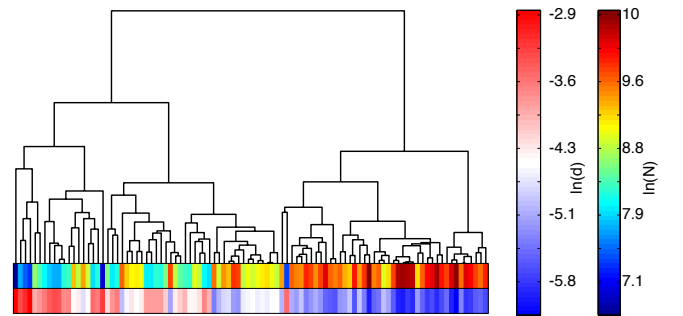


FIG. 11. (Color online) Dendrogram for 100 Facebook networks of US universities at a single-time snapshot in September 2005. We order the leaves of the dendrogram to minimize the distance between adjacent nodes. The color bars below the dendrogram indicate (top) the number of nodes N in each network and (bottom) the fraction of possible edges that are present d .

In Fig. 11, we show the dendrogram for Facebook networks that we produced by comparing MRFs. The two color bars below the dendrogram indicate (top) the number of nodes N in each network and (bottom) the frac-

tion of possible edges d that are present (i.e., edge density). The Facebook networks range in size from 762 to 41,536 nodes, and the edge density varies from 0.2% to 6%. In contrast to previous examples, we observe in this case that two simple network properties appear to explain most of the observed clustering of the networks. An important feature of this example is that the \mathcal{H}_{eff} , S_{eff} , and η_{eff} MRFs are each very similar in shape and lie in a narrow range across all 100 institutions (see Fig. 7). Such extreme similarity is remarkable—as one can see in Fig. 7, this contrasts starkly with most of the other examples—and it suggests that all of the Facebook networks have very similar mesoscopic structural features. If one also considers demographic information, then one can find interesting differences between the networks [21], but the structural similarity is striking.

D. Fungi

We also examined fungal mycelial networks extracted from time series of digitized images of colony growth. In these undirected, planar, weighted networks, the nodes represent hyphal tips, branch points, or anastomoses (hyphal fusions), and the edges represent the interconnecting hyphal cords weighted by their conductivity [27, 52, 53]. For comparison, we also digitized weighted networks of the acellular slime mold *Physarum polycephalum* [24]. Fungal networks look like trees but contain additional edges (known as *cross-links*) that generate cycles.

As shown in Fig. 12(a), we find using our method that replicate networks from different species at comparable time points are grouped together. Furthermore, the aggregate clustering pattern reflects increasing levels of cross-linking that are characteristic of different species, as illustrated in Fig. 12(b); this ranges from the low levels in *Resinicium bicolor* to intermediate levels in *Phanerochaete velutina* and highly cross-linked networks formed by *Phallus impudicus*. By constructing a dendrogram for only one species but including data from repeated experiments and over time (see Fig. 12(c)), we observe a progression from trees at early developmental times to an increasingly cross-linked network later in mycelium growth [26, 27]. In early growth, the developmental stage appears to dominate the clustering pattern, as networks from different replicates but of similar age are grouped together. At later times, however, networks show a high aggregate level of similarity, and the fine-grained clustering predominantly reflects the subtle changes in structure evolving within each replicate.

E. New York Stock Exchange

As our final example, we consider a set of stock-return correlation networks for the New York Stock Exchange (NYSE), which is the largest stock exchange in the world (as measured by the aggregate US dollar value of the

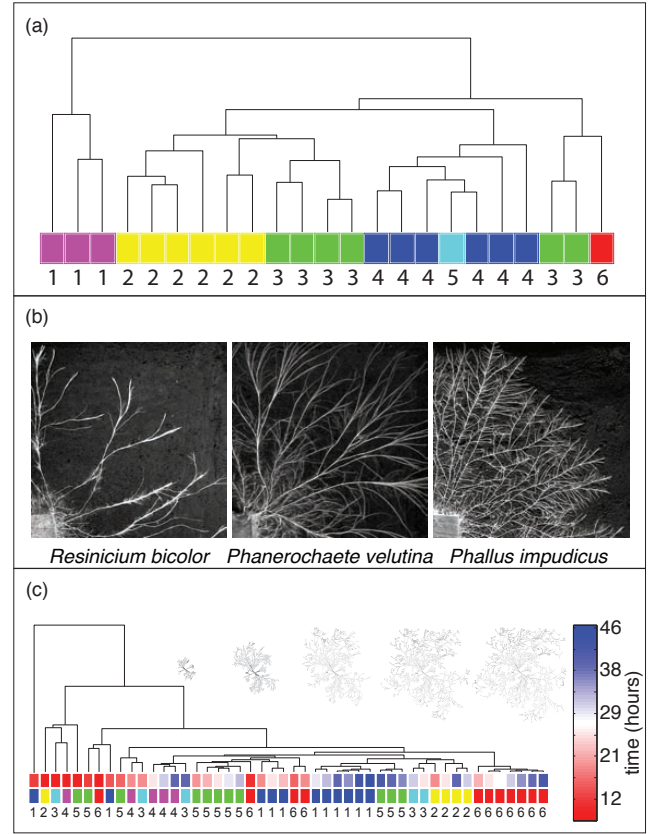


FIG. 12. (Color online) (a) Dendrogram of networks for six different species of *Saprotrophic basidiomycetes* and the slime mold *Physarum polycephalum*. Each leaf represents a replicate experiment. The colors and numbers correspond to the species as follows: (1) *Resinicium bicolor*, (2) *Physarum polycephalum*, (3) *Phallus impudicus*, (4) *Phanerochaete velutina*, (5) *Stropharia caerulea*, and (6) *Agrocybe gibberosa*. (b) Images illustrating the network structure of the different species [52]. (c) Dendrogram of network development in six replicate time series of *Phanerochaete velutina*. We color the leaves by time, and the color bar underneath the leaves indicates experiment number (1, ..., 6). In the inset, we show extracted networks that illustrate the transition from simple branching trees to increasing levels of interconnection (i.e., cross-linking) with time.

securities listed on it). Each node represents a stock, and the strength of the edge connecting stocks i and j is linear in the Pearson product-moment correlation coefficient between the daily logarithmic returns of the stocks [22]. We consider $N = 100$ stocks during the time period 1985–2008 and construct a network for each 6 months of data. This yields a sequence of fully-connected, weighted adjacency matrices whose elements quantify the similarity of two stocks (normalized to the unit interval for each time window).

We show the dendrogram for the NYSE networks in Fig. 13. The first division of these networks classifies them into two groups (which we have colored in blue and

red). The red cluster appears to correspond to periods of market turmoil, including the networks for the second half of 1987 (including the Black Monday crash of October 1987), all of 2000–2002 (including and following the bursting of the dot-com bubble), and the second half of 2007 and all of 2008 (including the recent credit and liquidity crisis). The value of the NYSE composite index, which measures the aggregate performance of all common stocks listed on the NYSE [56], supports our hypothesis that the red cluster is associated with periods of market turmoil. Indeed, the networks in the red cluster correspond (with one or two exceptions) to the periods of high volatility of the composite index (see Fig. 13).

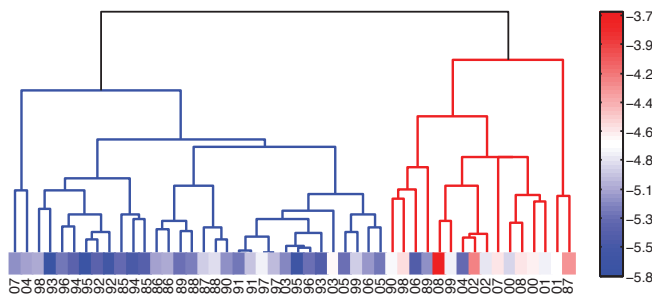


FIG. 13. (Color online) Dendrogram for 48 NYSE networks during the period 1985–2008 [22]. Observe the clear split of the dendrogram into two clusters (a blue group on the left and a red group on the right). Leaf color indicates mean daily volatility of the composite index.

X. CONCLUSIONS

We have developed an approach that facilitates the comparison of diverse networks by summarizing network community structure using what we call mesoscopic response functions (MRFs). We have demonstrated how this approach can be used to group networks both across categories and within categories. Our work builds on prior research on network community structure, which has focused predominantly on algorithmic detection of the communities rather than on subsequently using the communities for applications (such as comparing sets of networks).

The development of algorithmic methods to detect communities is frequently motivated by the idea that the community structure of a network representing a system has some bearing on the function of the system. If different networks perform different functions—and if their functions are constrained, at least in part, by their mesoscopic structure—then it should be possible in principle to derive a functional classification of networks based on community structure. Although this has mostly been presented as a presumption in the existing literature, it is actually an empirically testable hypothesis. Indeed,

we have shown in the present paper that one can systematically exploit mesoscopic structure to obtain useful comparisons of networks. This allows one to derive taxonomies for networks that also appear to have correspondence with functional similarities. We observed that networks that were not grouped with other members of the same class appeared to be unusual in some respects, and we also demonstrated that we could detect historically-noted financial and political changes from time-ordered sequences of networks.

We believe that our framework has the potential to aid in the exploration and exploitation of similarities in network structures across both network types and disciplinary boundaries.

ACKNOWLEDGMENTS

Each of the networks that we have used is described in detail in the Supplemental Material, where we have also provide relevant citations for the data sources. We are particularly grateful to A. D’Angelo and Facebook, A. Merdzanovic, M. E. J. Newman, E. Voeten, and Voteview [30] (maintained by K. Poole and H. Rosenthal) for data. We thank A. Lewis, G. Villar, S. Agarwal, D. Smith, and the anonymous referees for useful suggestions. We adapted code originally written by R. Guimerà for the simulated annealing algorithm and code by R. Lambiotte for the Louvain algorithm. This research was supported in part by the Fulbright Program (JPO) and NIH grant P01 AG-031093 (JPO), the James S. McDonnell Foundation (MAP: #220020177), NSF (PJM: DMS-0645369), and EPSRC and BBSRC (NSJ: EP/I005765/1, EP/I005986/1, and BBD0201901).

Appendix A: Robustness of Clustering

To examine the robustness of our clustering to false positives (false links) and false negatives (false non-links), we consider two network rewiring mechanisms, and we apply the rewiring to each network in a subset of 25 networks highlighted in Table 2 of the Supplemental Material. The first step in the procedure is to randomly rewire a number of edges corresponding to a given percentage (5%, 10%, 20%, 50%, or 100%) of the total number of edges in the network, subject to the constraints that we preserve the networks’s degree distribution and the fact that it consists of a single connected component [57]. (That is, such a rewiring of a number of edges equal to $x\%$ of the L edges in a network means that we perform $\lceil xL \rceil$ rewiring steps; the same edge can be rewired multiple times.) Second, we randomly rewire a given number of the edges subject only to the constraint that the rewired network still consists of a single component.

Because we are perturbing the original network, we focus on the distance matrices \mathbf{D}^H , \mathbf{D}^S , and \mathbf{D}^n as they can be calculated directly for each network. We consider 25

of the 746 original networks of varying sizes and edge densities; we highlight these networks in bold in Table II of the Supplemental Material. In Fig. 14, we show the distance matrices for this subset of networks when different percentages of edges have been rewired with the degree distribution preserved. The first column shows the matrices for the original networks. (Note that the node orderings for $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^{η} are not necessarily the same in Fig. 14 because of the block-diagonalization of matrices.) The subsequent columns show the mean distance matrices as increasing numbers of edges are rewired; for a given row, the node ordering in each column is fixed. The distance matrices for the randomizations are the mean pairwise distances between networks, where the mean is calculated over all possible pairs between 10 perturbations of each network. More precisely, let A and B represent two different (unperturbed) networks and let the sequences A_1, A_2, \dots, A_{10} and B_1, B_2, \dots, B_{10} represent 10 realizations of the perturbation process (e.g., at the 5% level) for the networks. To calculate the distance between A and B under perturbation, we find for each $j \in \{1, \dots, 10\}$ the distances between A_j and B_1 , A_j and B_2 , \dots , and A_j and B_{10} . We then calculate the mean of the ensuing $10 \times 10 = 100$ distance values. Based on visual inspection of Fig. 14, the matrices for the first few columns for all of the distances are fairly similar to the original distance matrices. This suggests some notion of robustness in our clustering technique. We study only 25 networks because of the computational costs of rewiring a large number of networks multiple times; however, we have performed the same investigation for 5 different subsets of 25 networks and obtained similar results. We list the networks in each subset of 25 in Table I in the Supplemental Material.

To carry out a more thorough randomization of each network, we now rewire every edge in the network 10 times on average. In Fig. 15, we show the $\mathbf{D}^{\mathcal{H}}$, \mathbf{D}^S , and \mathbf{D}^{η} mean-distance matrices for this number of rewirings. We again calculate the mean distance using the method described in the previous paragraph. The first column again shows the distance matrices for the original networks. The second and third columns show the distance matrices for randomizations in which the degree distribution is preserved and destroyed, respectively. The node orderings of the matrices in the second and third columns are again the same as the orderings for the matrix of the first column of the corresponding row. The second column in Fig. 15 demonstrates that some block structure remains in the distance matrices when the degree distribution is preserved. The third column shows that much of this structure is destroyed (though some block structure is still visible) when the degree distribution is not preserved. When the networks are “fully randomized” in this way—with the only constraint being that each rewired network must consist of a single connected component—one is in effect producing random graphs. These random graphs might, however, still have some common properties, such as the number of nodes and

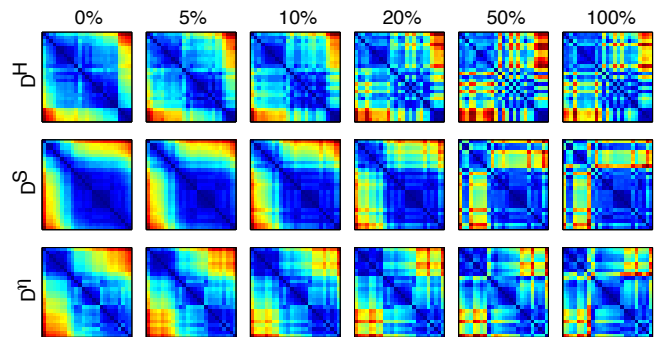


FIG. 14. (Color online) Block-diagonalized mean distance matrices $\mathbf{D}^{\mathcal{H}}$ (top row), \mathbf{D}^S (middle row), and \mathbf{D}^{η} (bottom row) for the 25 networks listed in bold in Table II of the Supplemental Material. The columns show the mean-distance matrices following randomizations of the original network in which a given percentage of edges are rewired and the degree distributions of the networks are preserved. (We also constrain each rewired network to consist of a single connected component.) The ordering of the nodes in the plots is fixed for each row. The first column shows the distance matrix for the original networks. The distance matrices for the randomizations are the mean pairwise distances between networks.

the edge density.

Appendix B: Computational Heuristics

1. Robustness of Network MRFs

We detected all communities in the main text using the locally greedy Louvain algorithm [25]; however, several alternative heuristics exist, so we now investigate whether the choice of heuristic has any effect on the results. In Ref. [24], Good et al. demonstrated that there can be extreme near-degeneracies in the energy function, in particular an exponential number of low-energy (i.e., high-modularity) solutions. Given this, it is unsurprising that different energy-optimization heuristics can yield very different partitions for the same network. Good et al. suggested that the reason for this behavior is that different heuristics sample different regions of the energy landscape. Because of the potential sensitivity of results to the choice of heuristic, one should treat individual partitions by particular heuristics with caution. However, one can have more confidence in the validity of the partitions if different heuristics produce similar results. Here we compare the results for the Louvain algorithm [25] with those for a spectral algorithm [18] and simulated annealing [58].

In Fig. 16, we show MRFs for three networks calculated using Louvain [25], spectral [18] and simulated annealing algorithms [58]. For all three networks, the three algorithms agree very closely on the shapes of the \mathcal{H} , S , and η MRFs. The MRFs are most similar for the roll-call vot-

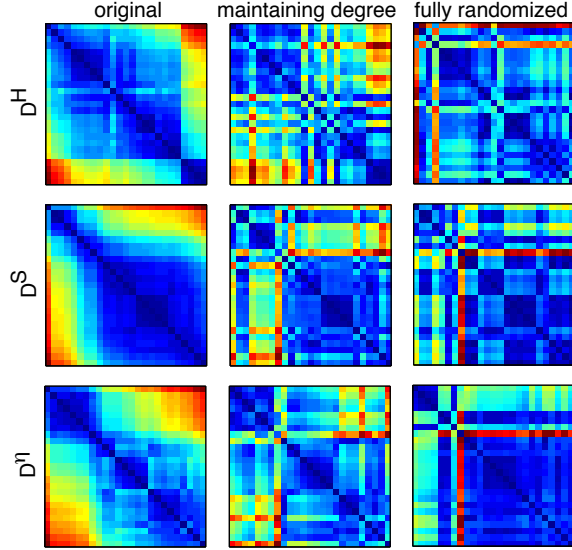


FIG. 15. (Color online) Block-diagonalized distance matrices \mathbf{D}^H (top row), \mathbf{D}^S (middle row), and \mathbf{D}^I (bottom row) for the 25 networks listed in bold in Table II of the Supplemental Material. The first column shows the distance matrices for the original networks. The second column shows the mean distance matrices following randomizations of the original networks in which 10 times the total number of edges in the networks have been rewired such that the degree distributions are preserved and the rewired networks each consist of a single connected component. The third column shows the mean distance matrices following randomizations of the original networks in which 10 times the total number of edges in the networks have been rewired but only the fact that the networks consist of single connected components is preserved (i.e., the degree distributions are not preserved). The distance matrices for the randomizations are composed of the mean pairwise distances between the networks.

ing network of the 102nd US Senate [50–52], and the \mathcal{H} MRF is almost identical for the three heuristics. In general, we observe the largest differences in the shapes of the MRFs when using the spectral algorithm. The spectral algorithm that we used begins by finding a partition of the network into exactly two components such that the energy is minimized (among all bipartitions). It then recursively partitions the smaller networks into two groups until no decrease in energy can be obtained through bipartitioning. At each step, this algorithm only finds the optimal partition of each community into two smaller communities even though a split into more communities could yield a lower energy. Given this, it is unsurprising that the spectral algorithm often identifies partitions further from the optimum than the other heuristics. For the remainder of this section, we therefore only compare the Louvain and simulated annealing algorithms.

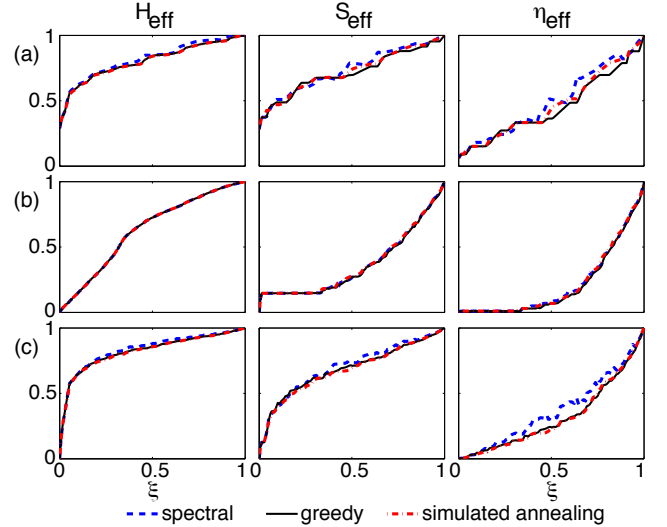


FIG. 16. (Color online) Comparison of the MRFs produced using spectral [18], Louvain [25], and simulated annealing [58] optimization heuristics. We show the MRFs for the (a) Zachary Karate Club network [83], (b) the roll-call voting network of the 102nd US Senate [50–52], and (c) the Garfield small-world citations network [40].

2. Robustness of Resulting Network Taxonomies

Although Fig. 16 shows good agreement between the shapes of the MRFs that we obtain from the different computational heuristics, we nevertheless check that the small differences that do occur do not have a significant effect on the resulting network taxonomy. Because of the computational cost of detecting communities using simulated annealing, we investigate the effect on the taxonomy using a subset of small networks. We highlight all of the networks that we consider with an asterisk (*) in Table II of the Supplemental Material. (The largest network that we include is the cat brain cortical/thalamic network [11], which has 1,170 nodes.). Indeed, MRFs for small networks tend to be much noisier than those for large networks—see, for example, Fig. 16(a), which shows the MRFs for the 34-node Zachary Karate Club network—so we expect that any differences between algorithms are likely to be more pronounced for small networks.

In Fig. 17, we show dendrograms obtained using the Louvain and simulated-annealing modularity optimization algorithms for a subset of 15 networks. On visual inspection, the dendrograms appear to be very similar, as there are only a few small differences in the heights at which leaves and clusters combine. To quantify the similarity between a pair of dendrograms with underlying distance matrices \mathbf{s} and \mathbf{t} , we define a correlation

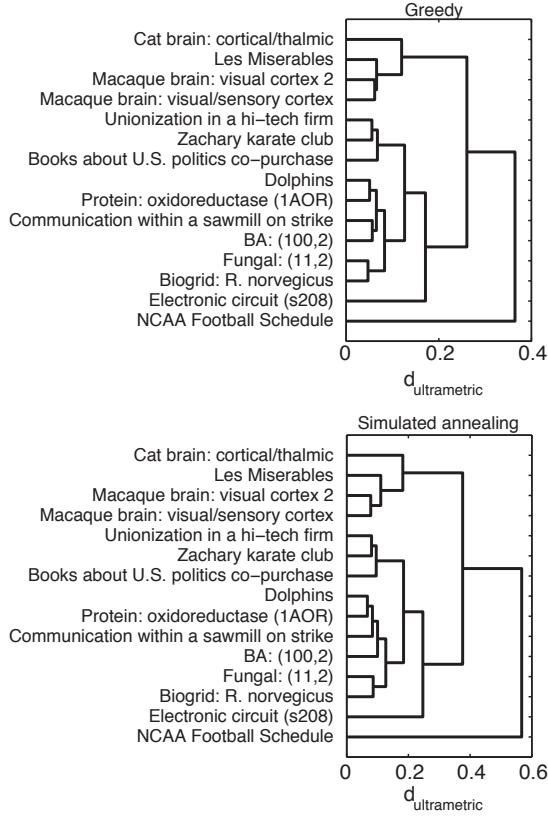


FIG. 17. (Color online) Comparison of the dendrograms produced using a Louvain algorithm (top panel) and simulated annealing (bottom panel) for a subset of 15 networks. The only difference between the two dendrograms is the order in which the “Communication within a sawmill on strike” and the “BA: (100,2)” networks cluster and the distances at which the other networks cluster.

coefficient φ as

$$\varphi = \frac{\sum_{i < j} (s_{ij} - \bar{s})(t_{ij} - \bar{t})}{\sqrt{\left[\sum_{i < j} (s_{ij} - \bar{s})^2 \right] \left[\sum_{i < j} (t_{ij} - \bar{t})^2 \right]}}, \quad (\text{B1})$$

where \bar{s} is the mean of the distances s_{ij} and \bar{t} is the mean of the distances t_{ij} . Dendrograms derived from identical distance matrices have correlation coefficient $\varphi = 1$. The correlation for the example dendrograms shown in Fig. 17 is 0.997. To get a better sense of the extent of this correlation, we compare the observed correlations with those obtained for randomized dendrograms. To make the comparison, we first produce a distribution of correlation coefficients φ between a large number of empirical (unrandomized) dendrograms produced by the Louvain and simulated-annealing algorithms. Because of the computational costs of calculating the MRFs for the simulated annealing algorithm, we only consider the subset of 25 networks identified above. We select 15 networks uniformly at random from this subset of 25 networks and

generate two dendrograms similar to those in Fig. 17: one corresponds to the distance matrix produced by the Louvain algorithm and the other corresponds to the distance matrix produced by simulated annealing. We then calculate the correlation coefficient between the two distance matrices. We repeat this process 10,000 times to obtain 10,000 correlation coefficients, whose distribution we show using the hollow red histogram in Fig. 18. This procedure makes it possible to compare a large number of dendrograms at the computational cost of calculating simulated annealing MRFs for a total of 25 networks, highlighted with asterisks in Table 2 of the Supplemental Material.

We then compare this observed distribution of correlation coefficients to a randomized reference. We focus on the correlation between empirical Louvain dendrograms (i.e., empirical dendrograms resulting from distance matrices produced by the Louvain method) and randomized simulated-annealing dendrograms (i.e., dendrograms resulting from distance matrices produced by the simulated annealing algorithm that have been subsequently randomized). We proceed as follows: for each of the 10,000 dendrogram pairs that we assembled from subsets of 15 networks, we create 100 randomizations of the simulated-annealing dendrogram, and we then calculate the correlation coefficient between each of these randomized dendrograms and the corresponding empirical Louvain dendrogram. The resulting distribution from 10,000 repetitions is the solid blue histogram in Fig. 18. To randomize the simulated-annealing dendrogram, we used the double-permutation procedure described in Refs. [60, 61]. This procedure has two steps. First, we randomize the distances at which the different clusters are combined. For example, consider an unrandomized dendrogram in which clusters A and B are combined at a distance of 0.45 and clusters C and D are combined at a distance of 0.65; after the randomization, A and B might be combined at a distance of 0.65 and C and D might be combined at a distance of 0.45. Second, we randomize the networks corresponding to each leaf in the dendrogram. This two-step randomization procedure maintains the underlying distances and the topology of the dendrogram.

As mentioned above, we show the distributions of correlation coefficients between empirical Louvain dendrograms and the empirical (unrandomized) and randomized simulated-annealing dendrograms in Fig. 18. The correlation is clearly much higher for the empirical case, as there is only a very slight overlap in the tails of the two distributions. The correlation between the Louvain and simulated-annealing dendrograms is greater than 0.99 for about 63% of the studied dendrograms.

Appendix C: Diagnostic for Assessing the Clustering from Different Distance Measures

An examination of the leaf colors of the dendrogram in Fig. 7 illustrates that the employed distance measure

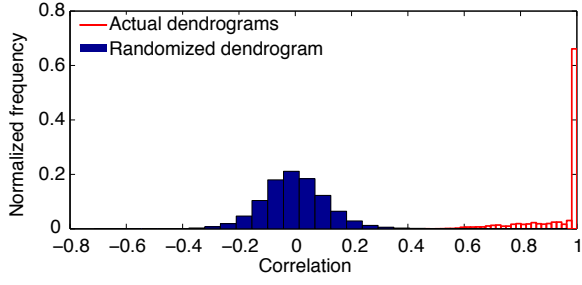


FIG. 18. (Color online) Comparison of the distributions of correlation coefficients between empirical Louvain dendrograms and empirical (red, hollow) and randomized (blue, solid) simulated-annealing dendrograms. See the text for details.

groups together networks from a variety of categories, including political voting networks, political committee networks, Facebook networks, metabolic networks, and fungal networks. A visual comparison provides a reasonable starting point for assessing the effectiveness of different distance measures at clustering networks. To quantify how effectively each distance matrix (\mathbf{D}^H , \mathbf{D}^S , \mathbf{D}^T , and \mathbf{D}^P) clusters networks of the same type, we introduce a *clustering diagnostic*, which we denote by $\alpha(h)$, to be explained shortly. Because the assignment of networks to categories is subjective and because some of the categories include networks of very different types, it would be inappropriate to assess the effectiveness of a distance measure based on how well it clusters networks in very broad categories. We thus focus our examination on narrower categories whose constituent networks are clustered fairly tightly in Fig. 7. This includes the following 8 categories of networks: Facebook, metabolic, political cosponsorship, political committee, political voting, financial, brain, and fungal.

The clustering diagnostic depends on where one “cuts” the dendrograms. We start by constructing a dendrogram for each of the four distance matrices \mathbf{D}^H , \mathbf{D}^S , \mathbf{D}^T , and \mathbf{D}^P . Performing a horizontal cut through a dendrogram at a given height h splits the dendrogram into multiple disconnected clusters (h is measured in terms of ultrametric distances; see Fig. 17). For each such cluster, we calculate the proportion of networks from a particular category that are contained in it. For example, if a cut produces three clusters and if we consider the Facebook category, then we might find that one cluster contains two tenths of the Facebook networks, a second cluster has three tenths of those networks, and the third cluster has the remaining half of those networks. We calculate these membership fractions for each network category and for each cluster. We then identify, for each category, what we called the *plurality cluster*, which is defined as the cluster that includes the largest fraction of networks from that category. In the above example, the third cluster is the plurality cluster for the Facebook category. Our diagnostic $\alpha(h)$ is then defined by adding across all 8 categories

the fraction of networks in the plurality clusters:

$$\alpha(h) = \sum_{j=1}^8 \gamma_j(h), \quad (\text{C1})$$

where $\gamma_j(h)$ is the plurality fraction for the j th category of networks for the given cut at height h of the dendrogram.

We perform similar calculations for each level of the dendrogram and use the resulting values of $\alpha(h)$ to assess the effectiveness of the different distance measures at clustering the networks. For example, at the root of the dendrogram, all of the networks are in a single cluster, so the maximum fraction of networks in the same cluster is 1 for every network category. Given the above choice of 8 categories, this yields $\alpha = 8$. However, as one considers lower levels of the dendrogram, the clusters break up more and more, so the fraction of networks in the plurality cluster in each category typically decreases. Effective distances measures ought to result in relatively high values for $\alpha(h)$.

In Fig. 19, we compare the values of $\alpha(h)$ at each level of the dendrogram for \mathbf{D}^H , \mathbf{D}^S , \mathbf{D}^T , and \mathbf{D}^P . For each of the different subsets of networks and for most of the dendrogram levels, the PCA-distance \mathbf{D}^P is the most effective of the employed distance measures at clustering networks of the same category. This agrees with our visual assessment (i.e., our identification of contiguous blocks of color) of the different measures.

Appendix D: Using Simple Characteristics to Cluster Networks

We established in Section VIII that the PCA-distances \mathbf{D}^P between MRFs can produce sensible network taxonomies, and we now consider briefly whether the observed taxonomies can be explained using simple summary statistics. We consider only a few specific properties, though of course there are myriad other network diagnostics that one might consider.

Perhaps the three simplest properties of an undirected network are the following: (1) whether it has weighted or unweighted edges; (2) the number of nodes N ; and (3) the edge density $d = 2L/[N(N-1)]$ (where L is the number of edges, which we distinguish from the total edge weight m in weighted networks). The top colored row in Fig. 20 indicates that many of the weighted networks are clustered together at the far left of the dendrogram. However, there are also weighted networks scattered throughout the dendrogram, so whether a network is weighted or unweighted does not explain the observed classification. The third colored row provides a clearer explanation for the cluster of networks at the left: These are not simply weighted networks, as they are in fact similarity networks, so that nearly all possible edges are present and have weights indicating connection strengths. However,

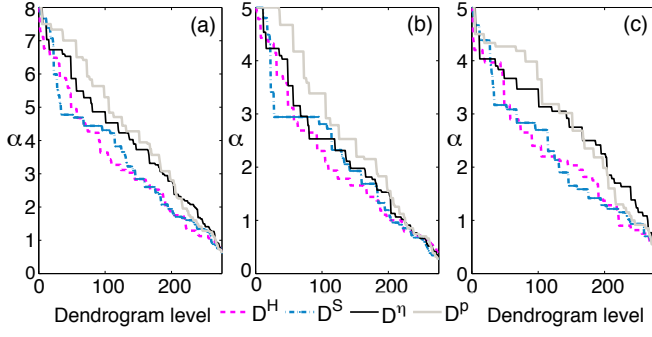


FIG. 19. (Color online) Comparison of the effectiveness of the employed distance measures at clustering networks of the same category. As discussed in this text, we quantify this using the clustering diagnostic $\alpha(h)$. We calculate dendrograms from four distance matrices (\mathbf{D}^H , \mathbf{D}^S , \mathbf{D}^η , and \mathbf{D}^P) and compare the resulting values of $\alpha(h)$ for different sets of categories. (a) The value of the clustering diagnostic $\alpha(h)$ as a function of dendrogram cut level h (i.e., where the dendrogram is split to clusters) for the following 8 categories of networks: Facebook, metabolic, political cosponsorship, political committee, political voting, financial, brain, and fungal. (b) The value of $\alpha(h)$ for the largest 5 of the above 8 categories (Facebook, metabolic, political cosponsorship, political committee, and political voting) and (c) for the smallest 5 of the above 8 categories (Facebook, metabolic, financial, brain, and fungal). The maximum possible value of $\alpha(h)$ in each panel is equal to the number of categories considered in each panel. The values of $\alpha(h)$ obtained using the PCA-distance matrix \mathbf{D}^P (gray solid curve) are usually higher than those obtained using the other three distance measures. This suggests that PCA distance is the most effective of the four employed clustering measures.

this property alone cannot explain the observed classification, as several of the weighted networks containing nearly all possible edges do not appear at the far left of the dendrogram. In fact, there are many clusters in the dendrogram that contain networks with very different fractions of possible edges. The total number of nodes, shown by the second colored row in the figure, again explains some of the clustering, as networks with similar numbers of nodes are clustered together in some regions of the dendrogram. However, there are also numerous examples in which networks with the same number of nodes appear in different clusters. Therefore, none of these three simple network diagnostics can explain the observed classification by itself.

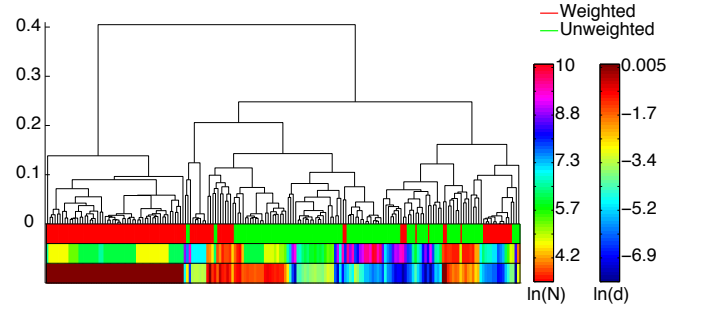


FIG. 20. (Color online) Taxonomy for 189 networks. We constructed the dendrogram using the distance matrix \mathbf{D}^P and average linkage clustering. We order the leaves of the dendrogram to minimize the distance between adjacent nodes, and we color the leaves to indicate the type of network. The three color bars below the dendrogram indicate whether the network corresponding to each leaf is weighted or unweighted (top), the number of nodes in the networks N (middle), and the fraction of possible edges that are present d (bottom).

-
- [1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, UK, 2010).
- [2] M. E. J. Newman, *Physics Today* **61**, 33 (2008).
- [3] K. Börner, *Atlas of Science: Visualizing What We Know* (MIT Press, Cambridge, MA, USA, 2010).
- [4] The code used for constructing network taxonomies in this paper is available at www.jponnela.com.
- [5] R. Albert and A.-L. Barabási, *Review of Modern Physics* **74**, 47 (2002).
- [37] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [32] R. Milo, S. Itzkovitz, N. Kashtan, R. Levittand, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [8] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, *Nature Physics* **3**, 63 (2007).
- [9] M. A. Serrano, M. Boguná, and A. Vespignani, *Proceedings of the National Academy of Sciences of the U.S.A.* **106**, 6483 (2009).
- [10] A. Arenas, J. Borge-Holthoefer, A. Gomez, and G. Zamora, *New Journal of Physics* **12**, 053009 (2010).
- [11] J. Bagrow, E. Bollt, J. Skufca, and D. Ben-Avraham, *EPL (Europhysics Letters)* **81**, 68004 (2008).
- [12] C. Song, S. Havlin, and H. A. Makse, *Nature* **433**, 392 (2005).
- [13] M. A. Porter, J.-P. Onnela, and P. J. Mucha, *Notices of the American Mathematical Society* **56**, 1082 (2009).
- [14] S. Fortunato, *Physics Reports* **486**, 75 (2010).
- [15] A. Lancichinetti, M. Kivela, J. Saramaki, and S. Fortunato, *PLoS One* **5**, e11976 (2010).
- [16] J. Reichardt and S. Bornholdt, *Physical Review E* **74**, 016110 (2006).
- [9] M. E. J. Newman and M. Girvan, *Physical Review E* **69**, 026113 (2004).
- [18] M. E. J. Newman, *Proceedings of the National Academy of Sciences of the U.S.A.* **103**, 8577 (2006).
- [19] J. Reichardt and S. Bornholdt, *Physical Review Letters* **93**, 218701 (2004).
- [21] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *SIAM Review* **53**, 526 (2011).
- [21] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *arXiv:1102.2166* (2011).
- [83] W. W. Zachary, *Journal of Anthropological Research* **33**, 452 (1977).
- [23] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hofer, Z. Nikoloski, and D. Wagner, *IEEE Transactions on Knowledge and Data Engineering* **20**, 172 (2008).
- [24] B. H. Good, Y. A. de Montjoye, and A. Clauset, *Physical Review E* **81**, 46106 (2010).
- [25] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, *Journal of Statistical Mechanics* **2008**, P10008 (2008).
- [22] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Physical Review E* **68**, 056110 (2003).
- [27] A. Arenas, A. Fernandez, and S. Gómez, *New Journal of Physics* **10**, 053039 (2008).
- [4] O. Sporns, *BioSystems* **85**, 55 (2006).
- [53] D. Firth and A. Spirling, Working paper (2003), available at <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic/firth/software/tapir/firth-spirling.pdf>.
- [30] K. T. Poole, (2009), Data available at <http://voteview.com>.
- [51] K. T. Poole and H. Rosenthal, *Congress: A Political-Economic History of Roll Call Voting* (Oxford University Press, Oxford, UK, 1997).
- [52] N. M. McCarty, K. T. Poole, and H. Rosenthal, *Polarized America: The Dance of Ideology and Unequal Riches* (MIT Press, Cambridge, MA, USA, 2007).
- [50] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter, *arXiv:0907.3509* (2010).
- [55] C. Stark, B. J. Breitzkreutz, T. Regul, L. Boucher, A. Breitzkreutz, and M. Tyers, *Nucleic Acids Research* **34**, D535 (2006).
- [40] E. Garfield, I. H. Sher, and R. J. Torpie, *The Use of Citation Data in Writing the History of Science* (The Institute for Scientific Information, Philadelphia, PA, USA, 1964).
- [1] P. Erdős and A. Rényi, *Publicationes Mathematicae* **6**, 290 (1959).
- [3] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [2] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [39] I. T. Jolliffe, *Principal Component Analysis* (Springer-Verlag, New York, NY, USA, 1986).
- [40] S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
- [41] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley, New York, NY, USA, 2001).
- [42] The three most common linkage clustering algorithms are single, average, and complete linkage clustering; they join clusters of objects based on the smallest, mean, and largest distance between objects in the clusters, respectively. We used the cophenetic correlation ζ to compare how well each algorithm preserves the pairwise distances between the networks [?]. We obtain the following cophenetic correlations for the different linkage clustering algorithms: $\zeta_{\text{single}} \doteq 0.65$, $\zeta_{\text{avg}} \doteq 0.78$, and $\zeta_{\text{comp}} \doteq 0.62$. Hence, dendrograms constructed using average linkage clustering preserve the distances in \mathbf{D}^p better than those constructed using the other two clustering techniques. We thus use average linkage clustering to construct all of the dendrograms in this paper.
- [20] P. Gleiser and L. Danon, *Advances in Complex Systems* **6**, 565 (2003).
- [44] L. Hakes, J. W. Pinney, D. L. Robertson, and S. C. Lovell, *Nature Biotechnology* **26**, 69 (2008).
- [45] H. Yu, P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, and et al., *Science* **322**, 104 (2008).
- [72] R. Alberich, J. Miro-Julia, and F. Rossello, *arXiv:cond-mat/0202174* (2002).
- [47] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Science* **328**, 876 (2010).
- [48] S. Kernell, G. C. Jacobson, and T. Kousser, *The Logic of American Politics*, 4th ed. (CQ Press, Washington, DC, USA, 2009).
- [49] K. T. Macon, P. J. Mucha, and M. A. Porter, *Physica A* **391**, 343 (2012).
- [50] We excluded the 19th session because only one resolution was considered by the UNGA that year.
- [27] D. P. Bebber, J. Hynes, P. R. Darrah, L. Boddy, and M. D. Fricker, *Proceedings of the Royal Society B* **274**,

- 2307 (2007).
- [52] M. Fricker, L. Boddy, and D. Bebbber, “Network organization of mycelial fungi,” in *The Mycota. Vol VIII, Biology of the Fungal Cell*, edited by R. J. Howard and N. A. R. Gow (Springer, Berlin, Germany, 2007) pp. 307–328.
 - [53] M. D. Fricker, L. Boddy, T. Nakagaki, and D. P. Bebbber, “Adaptive biological networks,” in *Adaptive Networks: Theory, Models and Applications*, edited by T. Gross and H. Sayama (Springer, Berlin/Heidelberg, Germany, 2009) pp. 51–70.
 - [24] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebbber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, *Science* **327**, 439 (2010).
 - [26] L. L. M. Heaton, E. López, P. K. Maini, M. D. Fricker, and N. S. Jones, *Proceedings of the Royal Society B* **277**, 3265 (2010).
 - [56] See <http://www.nyse.com/>.
 - [57] S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
 - [58] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, *Physical Review E* **70**, 025101 (2004).
 - [11] J. W. Scannell, G. A. P. C. Burns, C. C. Hilgetag, M. A. O’Neil, and M. P. Young, *Cerebral Cortex* **9**, 277 (1999).
 - [60] F.-J. Lapointe and P. Legendre, *Journal of Classification* **12**, 265 (1995).
 - [61] F.-J. Lapointe and P. Legendre, *Systematic Zoology* **39**, 1 (1990).

SUPPLEMENTAL MATERIAL

Taxonomy Of All Studied Networks

In Supplemental Fig. 1, we show a dendrogram containing leaves for all of the 746 networks that we studied. This dendrogram contains several large contiguous blocks of leaves that correspond to networks belonging to the same category. For example, there are large contiguous blocks of fungal, Facebook, metabolic, political committee, political voting, and financial networks. These blocks do not always include all of the networks within a category; when there are separate contiguous blocks for the same category, the blocks sometimes correspond to different types of networks within a category. For example, the political voting networks category includes separate blocks of UN voting networks and UK House of Commons voting networks. However, because of the number of networks that we include in the study and the imbalance in the spread of networks across categories, Fig. 1 is difficult to interpret and the smaller categories are obfuscated by the larger ones. Therefore, for a clearer view of the relationships between the different categories of networks, we focus in the main text on 189 of the 746 networks.

Details of Networks

In Supplemental Table II, we provide details of all of the networks that we employed in our study. The set of networks includes several synthetic network models as well as synthetic benchmark networks that were introduced to test community detection algorithms. We include multiple realizations (using various parameter values) for many of the model and benchmark networks. In this section, we briefly describe these synthetic networks and explain the notation that we use to label them in Supplemental Table II.

1. Erdős-Rényi (ER)

In an ER network of N nodes, each pair of nodes is connected by an unweighted edge with probability p (and is not connected with probability $1 - p$) [1]. The degree of each node is distributed according to a binomial distribution. We label the ER networks using the notation “ER: (N, p) ”.

2. Watts-Strogatz (WS)

We consider the small-world network of Watts and Strogatz [2] for a one-dimensional lattice of N nodes with periodic boundary conditions. The network consists of a ring in which each node is connected with an unweighted

edge to all of its neighbors that are k or fewer lattice spacings away. Each edge is then considered in turn and one end is rewired with probability p to a different node selected uniformly at random, subject to the constraint that there can be no self-edges or multi-edges. We label each Watts-Strogatz network as “WS: (N, k, p) ”.

3. Barabási-Albert (BA)

BA networks [3] are obtained using a network growth mechanism in which nodes with degree m are added to the network, one per time step, and the other end of each new edge attaches to an existing node with a probability proportional to the degree of that node. We label each BA network as “BA: (N, m) ”.

4. Fractal

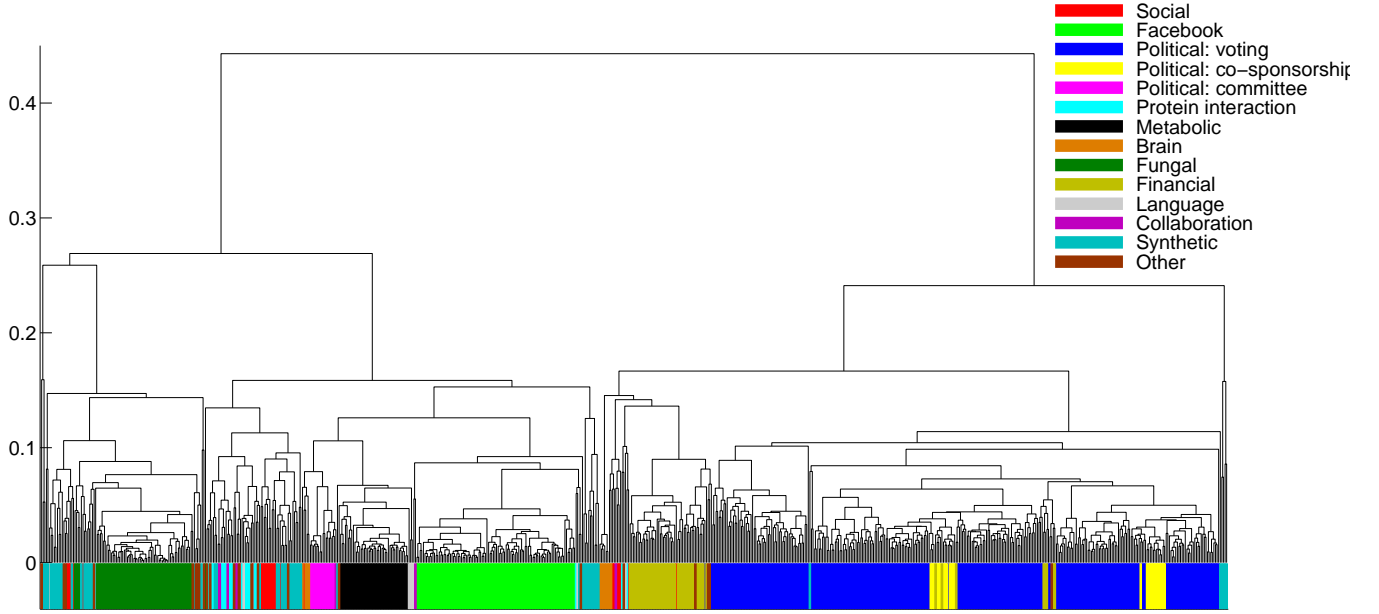
We generate fractal networks using the method described in Ref. [4]. We begin by generating an isolated group of 2^β fully connected nodes, where β gives the size of the clusters. These groups correspond to the hierarchical level $h = 0$. We then create a second identical group and connect the two groups using an edge density of f_e^{-1} , where f_e is the number of edges out of all possible edges between the groups. We then duplicate this network and connect the two duplicates at the level $h = 2$ using an edge density of f_e^{-2} . We repeat this until we reach the desired network size $N = 2^n$, where n is the number of hierarchical levels. At each step, the connection density is decreased, resulting in progressively sparser interconnectivity at higher hierarchical levels. The resulting network exhibits self-similar properties. We label each network “Fractal: (n, β, f_e) ”.

5. Random Fully-Connected

We produce randomly weighted, fully connected networks of N nodes by connecting every node to every other node with an edge whose weight is chosen uniformly at random on the unit interval. The networks have $N(N - 1)/2$ edges. We label each network “Random fully-connected: (N) ”.

6. Kumpula-Onnela-Saramäki-Kaski-Kertész (KOSKK) model

We generate weighted networks containing communities using the model described in Ref. [5]. We create edges via two mechanisms. First, at every time-step, each node i selects a neighbor j with probability w_{ij}/s_i , where w_{ij} is the weight of the edge connecting i and j and $s_i = \sum_j w_{ij}$ is the strength of i . If j has other neighbors



SM FIG. 1. (Color online) Dendrogram for 746 networks obtained using mesoscopic response functions (MRFs). Note that the colors used to indicate network categories are not the same as those in the main text.

in addition to i , then one of them is selected with probability $w_{jk}/(s_j - w_{ij})$. If i and k are not connected, then a new edge of weight $w_{ik} = w_0$ is created with probability p_n . If the edge already exists, its weight is increased by an amount δ . In both cases, w_{ij} and w_{jk} are also increased by δ . This process is termed *local attachment*. Second, if a node has no edges, then with probability p_r it creates an edge of weight w_0 to a randomly selected node. (This is called *global attachment*.) A node can be deleted with probability p_d ; if this happens, then all of its edges are also removed and the node is replaced by a new node, so that the total number of nodes N and the mean number of edges both remain constant. We label each network “Weighted: $(N, w_0, \delta, p_n, p_r, p_d, t)$ ”, where t is the total number of simulation time steps.

7. Lancichinetti-Fortunato-Radicchi (LFR) benchmark

The LFR benchmark [6] consists of unweighted networks with non-overlapping communities. A network in this ensemble is constructed by assigning each node a degree from a power-law distribution with exponent γ , where the extremes of the distribution k_{\min} and k_{\max} are chosen so that the mean degree is $\langle k \rangle$, and the nodes are connected using the configuration model [7]. Each node shares a fraction μ of its edges with nodes in other communities and a fraction $1 - \mu$ of them with nodes in its own community. The community sizes are taken from a power-law distribution with exponent β , subject to the

constraint that the sum of all of the community sizes equals the number of nodes N in the network. The minimum and maximum community sizes (q_{\min} and q_{\max}) are then chosen to satisfy the additional constraint that $q_{\min} > k_{\min}$ and $q_{\max} > k_{\max}$, which ensures that each node is included in at least one community. We label each network “LFR: $(N, \langle k \rangle, k_{\max}, \gamma, \beta, \mu, q_{\min}, q_{\max})$ ”.

8. Lancichinetti-Fortunato (LF) benchmark

The LF benchmark [8] allows networks to be weighted and the communities to overlap. In the present paper, we only consider weighted networks with non-overlapping communities. The node degrees are again taken from a power-law degree distribution (as in LFR benchmark networks), but this time we label the exponent τ_1 , and the community sizes are taken from a power-law degree distribution with exponent τ_2 . The strength s_i of each node is chosen so that $s_i = k_i^\beta$, where k_i again gives the degree of node i . There are also two mixing parameters: a topological mixing parameter μ_t , which specifies the proportion of edges outside a node’s community; and a participation mixing parameter μ_w , which specifies the weight of a node’s edges outside its community. We label each network “LF: $(N, \langle k \rangle, k_{\max}, \mu_t, \mu_w, \beta, \tau_1, \tau_2)$ ”. For all of the LF networks, we set $N = 1000$. One can alternatively set the minimum and maximum community sizes q_{\min} and q_{\max} . We always use $q_{\min} = 20$ and $q_{\max} = 50$, so we do not include these parameters when we label the networks.

9. LF-Newman-Girvan benchmark

We include an LF network ensemble with parameters values $N = 128$, $\langle k \rangle = 16$, $k_{\max} = 16$, $\mu_w = 0.1$, $q_{\min} = 32$, $q_{\max} = 32$, and $\beta = 1$. This family of networks is similar to the NG benchmark [8, 9].

SM TABLE I. The networks included in the 6 subsets of 25 networks used to test the robustness of the clusterings to random perturbations in Appendix A of the main text. The network ID corresponds to the numerical identifier of the network in Supplemental Table II.

Subset	Network IDs
1	9, 11, 243, 251, 267, 269, 280, 283, 301, 305, 340, 351, 353, 646, 662, 665, 674, 688, 690, 693, 700, 711, 735, 736, 740
2	71, 250, 251, 252, 264, 267, 270, 271, 276, 285, 301, 303, 305, 347, 354, 646, 649, 669, 674, 683, 688, 691, 737, 738, 739
3	11, 12, 243, 254, 264, 266, 268, 273, 284, 290, 291, 302, 308, 340, 341, 354, 355, 645, 649, 662, 672, 690, 695, 697, 717
4	252, 253, 258, 261, 264, 268, 274, 279, 280, 286, 288, 291, 342, 347, 348, 352, 641, 669, 694, 696, 699, 700, 711, 714, 737
5	10, 20, 34, 248, 250, 261, 266, 268, 272, 283, 285, 301, 306, 308, 344, 656, 661, 666, 693, 694, 700, 709, 711, 712, 715
6	34, 249, 252, 253, 256, 257, 266, 277, 283, 289, 291, 301, 302, 340, 341, 345, 650, 656, 661, 662, 690, 709, 713, 717, 736

SM TABLE II. Network summary statistics. We symmetrize all networks, remove self-edges, and only consider largest connected components. In this table, we give the network category, whether it is weighted or unweighted, the number of nodes N in the largest connected component, the number of edges L in this component, the fraction of possible edges present $f_e = 2L/[N(N-1)]$, and a reference providing details of the data source. We highlight the 25 networks used in the randomizations in Appendix A in bold and the 189 networks used in the aggregate taxonomy in red. We indicate with an asterisk (*) all networks used in Appendix B.2 to test the robustness of the taxonomy to different optimization heuristics.

ID	Name	Category	Weighted	N	L	f_e	References
1	Human brain cortex: participant A1	Brain	Y	994	13,520	0.0274	[10]
2	Human brain cortex: participant A2	Brain	Y	987	14,865	0.0305	[10]
3	Human brain cortex: participant B	Brain	Y	980	14,222	0.0296	[10]
4	Human brain cortex: participant D	Brain	Y	996	14,851	0.0300	[10]
5	Human brain cortex: participant E	Brain	Y	992	14,372	0.0292	[10]
6	Human brain cortex: participant C	Brain	Y	996	14,933	0.0301	[10]
7	Cat brain: cortical*	Brain	Y	52	515	0.3884	[11]
8	Cat brain: cortical/thalamic*	Brain	Y	95	1,170	0.2620	[11]
9	Macaque brain: cortical*	Brain	N	47	313	0.2895	[12]
10	Macaque brain: visual/sensory cortex*	Brain	N	71	438	0.1763	[12]
11	Macaque brain: visual cortex 1*	Brain	N	30	190	0.4368	[13]
12	Macaque brain: visual cortex 2*	Brain	N	32	194	0.3911	[13]
13	Coauthorship: astrophysics	Collaboration	Y	14,845	119,652	0.0011	[14]
14	Coauthorship: comp. geometry	Collaboration	Y	3,621	9,461	0.0014	[15, 16]
15	Coauthorship: condensed matter	Collaboration	Y	13,861	44,619	0.0005	[14]
16	Coauthorship: Erdős	Collaboration	N	6,927	11,850	0.0005	[17]
17	Coauthorship: high-energy theory	Collaboration	Y	5,835	13,815	0.0008	[14]
18	Coauthorship: network science	Collaboration	Y	379	914	0.0128	[18]
19	Hollywood film music*	Collaboration	Y	39	219	0.2955	[19]
20	Jazz collaboration	Collaboration	N	198	2,742	0.1406	[20]
21	Facebook: American	Facebook	N	6,370	217,654	0.0107	[21]
22	Facebook: Amherst	Facebook	N	2,235	90,954	0.0364	[21]
23	Facebook: Auburn	Facebook	N	18,448	973,918	0.0057	[21]
24	Facebook: Baylor	Facebook	N	12,799	679,815	0.0083	[21]
25	Facebook: BC	Facebook	N	11,498	486,961	0.0074	[21]
26	Facebook: Berkeley	Facebook	N	22,900	852,419	0.0033	[21]
27	Facebook: Bingham	Facebook	N	10,001	362,892	0.0073	[21]
28	Facebook: Bowdoin	Facebook	N	2,250	84,386	0.0334	[21]
29	Facebook: Brandeis	Facebook	N	3,887	137,561	0.0182	[21]
30	Facebook: Brown	Facebook	N	8,586	384,519	0.0104	[21]
31	Facebook: BU	Facebook	N	19,666	637,509	0.0033	[21]
32	Facebook: Bucknell	Facebook	N	3,824	158,863	0.0217	[21]
33	Facebook: Cal	Facebook	N	11,243	351,356	0.0056	[21]
34	Facebook: Caltech	Facebook	N	762	16,651	0.0574	[21]
35	Facebook: Carnegie	Facebook	N	6,621	249,959	0.0114	[21]
36	Facebook: Colgate	Facebook	N	3,482	155,043	0.0256	[21]
37	Facebook: Columbia	Facebook	N	11,706	444,295	0.0065	[21]
38	Facebook: Cornell	Facebook	N	18,621	790,753	0.0046	[21]
39	Facebook: Dartmouth	Facebook	N	7,677	304,065	0.0103	[21]
40	Facebook: Duke	Facebook	N	9,885	506,437	0.0104	[21]
41	Facebook: Emory	Facebook	N	7,449	330,008	0.0119	[21]
42	Facebook: FSU	Facebook	N	27,731	1,034,799	0.0027	[21]
43	Facebook: Georgetown	Facebook	N	9,388	425,619	0.0097	[21]
44	Facebook: GWU	Facebook	N	12,164	469,511	0.0063	[21]
45	Facebook: Hamilton	Facebook	N	2,312	96,393	0.0361	[21]
46	Facebook: Harvard	Facebook	N	15,086	824,595	0.0072	[21]
47	Facebook: Haverford	Facebook	N	1,446	59,589	0.0570	[21]
48	Facebook: Howard	Facebook	N	4,047	204,850	0.0250	[21]
49	Facebook: Indiana	Facebook	N	29,732	1,305,757	0.0030	[21]
50	Facebook: JMU	Facebook	N	14,070	485,564	0.0049	[21]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
51	Facebook: Johns	Facebook	N	5,157	186,572	0.0140	[21]
52	Facebook: Lehigh	Facebook	N	5,073	198,346	0.0154	[21]
53	Facebook: Maine	Facebook	N	9,065	243,245	0.0059	[21]
54	Facebook: Maryland	Facebook	N	20,829	744,832	0.0034	[21]
55	Facebook: Mich	Facebook	N	3,745	81,901	0.0117	[21]
56	Facebook: Michigan	Facebook	N	30,106	1,176,489	0.0026	[21]
57	Facebook: Middlebury	Facebook	N	3,069	124,607	0.0265	[21]
58	Facebook: Mississippi	Facebook	N	10,519	610,910	0.0110	[21]
59	Facebook: MIT	Facebook	N	6,402	251,230	0.0123	[21]
60	Facebook: MSU	Facebook	N	32,361	1,118,767	0.0021	[21]
61	Facebook: MU	Facebook	N	15,425	649,441	0.0055	[21]
62	Facebook: Northeastern	Facebook	N	13,868	381,919	0.0040	[21]
63	Facebook: Northwestern	Facebook	N	10,537	488,318	0.0088	[21]
64	Facebook: Notre Dame	Facebook	N	12,149	541,336	0.0073	[21]
65	Facebook: NYU	Facebook	Y	21,623	715,673	0.0031	[21]
66	Facebook: Oberlin	Facebook	N	2,920	89,912	0.0211	[21]
67	Facebook: Oklahoma	Facebook	N	17,420	892,524	0.0059	[21]
68	Facebook: Penn	Facebook	N	41,536	1,362,220	0.0016	[21]
69	Facebook: Pepperdine	Facebook	N	3,440	152,003	0.0257	[21]
70	Facebook: Princeton	Facebook	N	6,575	293,307	0.0136	[21]
71	Facebook: Reed	Facebook	N	962	18,812	0.0407	[21]
72	Facebook: Rice	Facebook	N	4,083	184,826	0.0222	[21]
73	Facebook: Rochester	Facebook	N	4,561	161,403	0.0155	[21]
74	Facebook: Rutgers	Facebook	N	24,568	784,596	0.0026	[21]
75	Facebook: Santa	Facebook	N	3,578	151,747	0.0237	[21]
76	Facebook: Simmons	Facebook	N	1,510	32,984	0.0290	[21]
77	Facebook: Smith	Facebook	N	2,970	97,133	0.0220	[21]
78	Facebook: Stanford	Facebook	N	11,586	568,309	0.0085	[21]
79	Facebook: Swarthmore	Facebook	N	1,657	61,049	0.0445	[21]
80	Facebook: Syracuse	Facebook	N	13,640	543,975	0.0058	[21]
81	Facebook: Temple	Facebook	N	13,653	360,774	0.0039	[21]
82	Facebook: Tennessee	Facebook	N	16,977	770,658	0.0053	[21]
83	Facebook: Texas80	Facebook	N	31,538	1,219,639	0.0025	[21]
84	Facebook: Texas84	Facebook	N	36,364	1,590,651	0.0024	[21]
85	Facebook: Trinity	Facebook	N	2,613	111,996	0.0328	[21]
86	Facebook: Tufts	Facebook	N	6,672	249,722	0.0112	[21]
87	Facebook: Tulane	Facebook	N	7,740	283,912	0.0095	[21]
88	Facebook: U. Chicago	Facebook	N	6,561	208,088	0.0097	[21]
89	Facebook: U. Conn.	Facebook	N	17,206	604,867	0.0041	[21]
90	Facebook: U. Illinois	Facebook	N	30,795	1,264,421	0.0027	[21]
91	Facebook: U. Mass.	Facebook	N	16,502	519,376	0.0038	[21]
92	Facebook: U. Penn.	Facebook	N	14,888	686,485	0.0062	[21]
93	Facebook: UC33	Facebook	N	16,800	522,141	0.0037	[21]
94	Facebook: UC61	Facebook	N	13,736	442,169	0.0047	[21]
95	Facebook: UC64	Facebook	N	6,810	155,320	0.0067	[21]
96	Facebook: UCF	Facebook	N	14,936	428,987	0.0038	[21]
97	Facebook: UCLA	Facebook	N	20,453	747,604	0.0036	[21]
98	Facebook: UCSB	Facebook	N	14,917	482,215	0.0043	[21]
99	Facebook: UCSC	Facebook	N	8,979	224,578	0.0056	[21]
100	Facebook: UCSD	Facebook	N	14,936	443,215	0.0040	[21]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
101	Facebook: UF	Facebook	N	35,111	1,465,654	0.0024	[21]
102	Facebook: UGA	Facebook	N	24,380	1,174,051	0.0040	[21]
103	Facebook: UNC	Facebook	N	18,158	766,796	0.0047	[21]
104	Facebook: USC	Facebook	N	17,440	801,851	0.0053	[21]
105	Facebook: USF	Facebook	N	13,367	321,209	0.0036	[21]
106	Facebook: USFCA	Facebook	N	2,672	65,244	0.0183	[21]
107	Facebook: UVA	Facebook	N	17,178	789,308	0.0054	[21]
108	Facebook: Vanderbilt	Facebook	N	8,063	427,829	0.0132	[21]
109	Facebook: Vassar	Facebook	N	3,068	119,161	0.0253	[21]
110	Facebook: Vermont	Facebook	N	7,322	191,220	0.0071	[21]
111	Facebook: Villanova	Facebook	N	7,755	314,980	0.0105	[21]
112	Facebook: Virginia	Facebook	N	21,319	698,175	0.0031	[21]
113	Facebook: Wake	Facebook	N	5,366	279,186	0.0194	[21]
114	Facebook: Wash. U.	Facebook	N	7,730	367,526	0.0123	[21]
115	Facebook: Wellesley	Facebook	N	2,970	94,899	0.0215	[21]
116	Facebook: Wesleyan	Facebook	N	3,591	138,034	0.0214	[21]
117	Facebook: William	Facebook	N	6,472	266,378	0.0127	[21]
118	Facebook: Williams	Facebook	N	2,788	112,985	0.0291	[21]
119	Facebook: Wisconsin	Facebook	N	23,831	835,946	0.0029	[21]
120	Facebook: Yale	Facebook	N	8,561	405,440	0.0111	[21]
121	NYSE: 1980-1999	Financial	Y	477	113,526	1.0000	[22]
122	NYSE: 1980-1983	Financial	Y	477	113,526	1.0000	[22]
123	NYSE: 1984-1987	Financial	Y	477	113,526	1.0000	[22]
124	NYSE: 1988-1991	Financial	Y	477	113,526	1.0000	[22]
125	NYSE: 1992-1995	Financial	Y	477	113,526	1.0000	[22]
126	NYSE: 1996-1999	Financial	Y	477	113,526	1.0000	[22]
127	NYSE: H1 1985	Financial	Y	100	4,950	1.0000	[23]
128	NYSE: H2 1985	Financial	Y	100	4,950	1.0000	[23]
129	NYSE: H1 1986	Financial	Y	100	4,950	1.0000	[23]
130	NYSE: H2 1986	Financial	Y	100	4,950	1.0000	[23]
131	NYSE: H1 1987	Financial	Y	100	4,950	1.0000	[23]
132	NYSE: H2 1987	Financial	Y	100	4,950	1.0000	[23]
133	NYSE: H1 1988	Financial	Y	100	4,950	1.0000	[23]
134	NYSE: H2 1988	Financial	Y	100	4,950	1.0000	[23]
135	NYSE: H1 1989	Financial	Y	100	4,950	1.0000	[23]
136	NYSE: H2 1989	Financial	Y	100	4,950	1.0000	[23]
137	NYSE: H1 1990	Financial	Y	100	4,950	1.0000	[23]
138	NYSE: H2 1990	Financial	Y	100	4,950	1.0000	[23]
139	NYSE: H1 1991	Financial	Y	100	4,950	1.0000	[23]
140	NYSE: H2 1991	Financial	Y	100	4,950	1.0000	[23]
141	NYSE: H1 1992	Financial	Y	100	4,950	1.0000	[23]
142	NYSE: H2 1992	Financial	Y	100	4,950	1.0000	[23]
143	NYSE: H1 1993	Financial	Y	100	4,950	1.0000	[23]
144	NYSE: H2 1993	Financial	Y	100	4,950	1.0000	[23]
145	NYSE: H1 1994	Financial	Y	100	4,950	1.0000	[23]
146	NYSE: H2 1994	Financial	Y	100	4,950	1.0000	[23]
147	NYSE: H1 1995	Financial	Y	100	4,950	1.0000	[23]
148	NYSE: H2 1995	Financial	Y	100	4,950	1.0000	[23]
149	NYSE: H1 1996	Financial	Y	100	4,950	1.0000	[23]
150	NYSE: H2 1996	Financial	Y	100	4,950	1.0000	[23]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
151	NYSE: H1 1997	Financial	Y	100	4,950	1.0000	[23]
152	NYSE: H2 1997	Financial	Y	100	4,950	1.0000	[23]
153	NYSE: H1 1998	Financial	Y	100	4,950	1.0000	[23]
154	NYSE: H2 1998	Financial	Y	100	4,950	1.0000	[23]
155	NYSE: H1 1999	Financial	Y	100	4,950	1.0000	[23]
156	NYSE: H2 1999	Financial	Y	100	4,950	1.0000	[23]
157	NYSE: H1 2000	Financial	Y	100	4,950	1.0000	[23]
158	NYSE: H2 2000	Financial	Y	100	4,950	1.0000	[23]
159	NYSE: H1 2001	Financial	Y	100	4,950	1.0000	[23]
160	NYSE: H2 2001	Financial	Y	100	4,950	1.0000	[23]
161	NYSE: H1 2002	Financial	Y	100	4,950	1.0000	[23]
162	NYSE: H2 2002	Financial	Y	100	4,950	1.0000	[23]
163	NYSE: H1 2003	Financial	Y	100	4,950	1.0000	[23]
164	NYSE: H2 2003	Financial	Y	100	4,950	1.0000	[23]
165	NYSE: H1 2004	Financial	Y	100	4,950	1.0000	[23]
166	NYSE: H2 2004	Financial	Y	100	4,950	1.0000	[23]
167	NYSE: H1 2005	Financial	Y	100	4,950	1.0000	[23]
168	NYSE: H2 2005	Financial	Y	100	4,950	1.0000	[23]
169	NYSE: H1 2006	Financial	Y	100	4,950	1.0000	[23]
170	NYSE: H2 2006	Financial	Y	100	4,950	1.0000	[23]
171	NYSE: H1 2007	Financial	Y	100	4,950	1.0000	[23]
172	NYSE: H2 2008	Financial	Y	100	4,950	1.0000	[23]
173	NYSE: H1 2008	Financial	Y	100	4,950	1.0000	[23]
174	NYSE: H2 2000	Financial	Y	100	4,950	1.0000	[23]
175	<i>Physarum polycephalum</i> 0126-bm06-wt-k2-1	Fungal	Y	411	645	0.0077	[24]
176	<i>Physarum polycephalum</i> 0149-bm05-wt-k2-1	Fungal	Y	345	548	0.0092	[24]
177	<i>Physarum polycephalum</i> 0157-bm03-wt-k2-1	Fungal	Y	251	399	0.0127	[24]
178	<i>Physarum polycephalum</i> 0166-bm03-wt-k2-1	Fungal	Y	492	778	0.0064	[24]
179	<i>Physarum polycephalum</i> 0181-bm02-wt-k2-1	Fungal	Y	192	307	0.0167	[24]
180	<i>Physarum polycephalum</i> 0185-bm02-wt-k2-1	Fungal	Y	199	311	0.0158	[24]
181	<i>Agrocybe gibberosa</i> AG-1	Fungal	Y	2366	3665	0.0013	[25]
182	<i>Phallus impudicus</i> PI113-1	Fungal	Y	543	725	0.0049	[25]
183	<i>Phallus impudicus</i> PI120-1	Fungal	Y	483	559	0.0048	[25]
184	<i>Phallus impudicus</i> PI37-1	Fungal	Y	644	826	0.0040	[25]
185	<i>Phallus impudicus</i> PI40-1	Fungal	Y	550	748	0.0050	[25]
186	<i>Phallus impudicus</i> PI-1	Fungal	Y	1,357	1,858	0.0020	[25]
187	<i>Resinicium bicolor</i> RB3ctl-3	Fungal	Y	202	233	0.0115	[25]
188	<i>Resinicium bicolor</i> RB4ctl-3	Fungal	Y	426	545	0.0060	[25]
189	<i>Resinicium bicolor</i> RB7ctl-3	Fungal	Y	380	458	0.0064	[25]
190	<i>Strophularia caerulea</i> SC-1	Fungal	Y	536	689	0.0048	[25]
191	<i>Phanerochaete velutina</i> control11-1	Fungal	Y	65	71	0.0341	[26]
192	<i>Phanerochaete velutina</i> control11-2	Fungal	Y	117	136	0.0200	[26]
193	<i>Phanerochaete velutina</i> control11-3	Fungal	Y	240	273	0.0095	[26]
194	<i>Phanerochaete velutina</i> control11-4	Fungal	Y	403	458	0.0057	[26]
195	<i>Phanerochaete velutina</i> control11-5	Fungal	Y	526	588	0.0043	[26]
196	<i>Phanerochaete velutina</i> control11-6	Fungal	Y	591	661	0.0038	[26]
197	<i>Phanerochaete velutina</i> control11-7	Fungal	Y	690	772	0.0032	[26]
198	<i>Phanerochaete velutina</i> control11-8	Fungal	Y	721	821	0.0032	[26]
199	<i>Phanerochaete velutina</i> control11-9	Fungal	Y	772	884	0.0030	[26]
200	<i>Phanerochaete velutina</i> control11-10	Fungal	Y	789	907	0.0029	[26]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
201	<i>Phanerochaete velutina</i> control11-11*	Fungal	Y	823	954	0.0028	[26]
202	<i>Phanerochaete velutina</i> control17-1	Fungal	Y	16	15	0.1250	[26]
203	<i>Phanerochaete velutina</i> control17-2	Fungal	Y	232	240	0.0090	[26]
204	<i>Phanerochaete velutina</i> control17-3	Fungal	Y	502	539	0.0043	[26]
205	<i>Phanerochaete velutina</i> control17-4	Fungal	Y	703	754	0.0031	[26]
206	<i>Phanerochaete velutina</i> control17-5	Fungal	Y	816	874	0.0026	[26]
207	<i>Phanerochaete velutina</i> control17-6	Fungal	Y	950	1,058	0.0023	[26]
208	<i>Phanerochaete velutina</i> control17-7	Fungal	Y	1,047	1,182	0.0022	[26]
209	<i>Phanerochaete velutina</i> control17-8	Fungal	Y	1,113	1,303	0.0021	[26]
210	<i>Phanerochaete velutina</i> control17-9	Fungal	Y	1,142	1,347	0.0021	[26]
211	<i>Phanerochaete velutina</i> control17-10	Fungal	Y	1,160	1,384	0.0021	[26]
212	<i>Phanerochaete velutina</i> control17-11	Fungal	Y	1,205	1,469	0.0020	[26]
213	<i>Phanerochaete velutina</i> control4-1	Fungal	Y	200	213	0.0107	[26]
214	<i>Phanerochaete velutina</i> control4-2	Fungal	Y	461	490	0.0046	[26]
215	<i>Phanerochaete velutina</i> control4-3	Fungal	Y	826	862	0.0025	[26]
216	<i>Phanerochaete velutina</i> control4-4	Fungal	Y	1,044	1,087	0.0020	[26]
217	<i>Phanerochaete velutina</i> control4-5	Fungal	Y	1,380	1,476	0.0016	[26]
218	<i>Phanerochaete velutina</i> control4-6	Fungal	Y	1,623	1,767	0.0013	[26]
219	<i>Phanerochaete velutina</i> control4-7	Fungal	Y	1,756	1,923	0.0012	[26]
220	<i>Phanerochaete velutina</i> control4-8	Fungal	Y	1,869	2,061	0.0012	[26]
221	<i>Phanerochaete velutina</i> control4-9	Fungal	Y	1,992	2,196	0.0011	[26]
222	<i>Phanerochaete velutina</i> control4-10	Fungal	Y	2,086	2,301	0.0011	[26]
223	<i>Phanerochaete velutina</i> control4-11	Fungal	Y	2,190	2,431	0.0010	[26]
224	<i>Phallus impudicus</i> pi150ctl-1	Fungal	Y	1,810	2,537	0.0015	[25]
225	<i>Phanerochaete velutina</i> pv81-1	Fungal	Y	75	82	0.0295	[27]
226	<i>Phanerochaete velutina</i> pv81-2	Fungal	Y	653	897	0.0042	[27]
227	<i>Phanerochaete velutina</i> pv81-3	Fungal	Y	911	1,255	0.0030	[27]
228	<i>Phanerochaete velutina</i> pv81-4	Fungal	Y	1,064	1,467	0.0026	[27]
229	<i>Phanerochaete velutina</i> pv81-5	Fungal	Y	986	1,351	0.0028	[27]
230	<i>Phanerochaete velutina</i> pv82-1	Fungal	Y	111	112	0.0183	[27]
231	<i>Phanerochaete velutina</i> pv82-2	Fungal	Y	467	523	0.0048	[27]
232	<i>Phanerochaete velutina</i> pv82-3	Fungal	Y	630	726	0.0037	[27]
233	<i>Phanerochaete velutina</i> pv82-4	Fungal	Y	644	749	0.0036	[27]
234	<i>Phanerochaete velutina</i> pv82-5	Fungal	Y	551	627	0.0041	[27]
235	<i>Phanerochaete velutina</i> pv83-1	Fungal	Y	129	142	0.0172	[27]
236	<i>Phanerochaete velutina</i> pv83-2	Fungal	Y	424	510	0.0057	[27]
237	<i>Phanerochaete velutina</i> pv83-3	Fungal	Y	671	857	0.0038	[27]
238	<i>Phanerochaete velutina</i> pv83-4	Fungal	Y	708	905	0.0036	[27]
239	<i>Phanerochaete velutina</i> pv83-5	Fungal	Y	551	668	0.0044	[27]
240	Online Dictionary of Computing	Language	Y	13,356	91,471	0.0010	[28]
241	Online Dictionary Of Information Science	Language	Y	2,898	16,376	0.0039	[16, 29]
242	Reuters 9/11 news	Language	Y	13,308	148,035	0.0017	[30]
243	Roget's thesaurus	Language	N	994	3,640	0.0074	[16, 31]
244	Word adjacency: English	Language	N	7,377	44,205	0.0016	[32]
245	Word adjacency: French	Language	N	8,308	23,832	0.0007	[32]
246	Word adjacency: Japanese	Language	N	2,698	7,995	0.0022	[32]
247	Word adjacency: Spanish	Language	N	11,558	43,050	0.0006	[32]
248	Metabolic: AA	Metabolic	N	411	1,818	0.0216	[33]
249	Metabolic: AB	Metabolic	N	386	1,691	0.0228	[33]
250	Metabolic: AG	Metabolic	N	494	2,173	0.0178	[33]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
251	Metabolic: AP	Metabolic	N	201	857	0.0426	[33]
252	Metabolic: AT	Metabolic	N	296	1,231	0.0282	[33]
253	Metabolic: BB	Metabolic	N	175	628	0.0412	[33]
254	Metabolic: BS	Metabolic	N	772	3,611	0.0121	[33]
255	Metabolic: CA	Metabolic	N	483	2,274	0.0195	[33]
256	Metabolic: CE	Metabolic	N	453	2,025	0.0198	[33]
257	Metabolic: CJ	Metabolic	N	370	1,631	0.0239	[33]
258	Metabolic: CL	Metabolic	N	382	1,646	0.0226	[33]
259	Metabolic: CQ	Metabolic	N	187	663	0.0381	[33]
260	Metabolic: CT	Metabolic	N	211	772	0.0348	[33]
261	Metabolic: CY	Metabolic	N	537	2,503	0.0174	[33]
262	Metabolic: DR	Metabolic	N	800	3,789	0.0119	[33]
263	Metabolic: EC	Metabolic	N	762	3,683	0.0127	[33]
264	Metabolic: EF	Metabolic	N	375	1,721	0.0245	[33]
265	Metabolic: EN	Metabolic	N	374	1,617	0.0232	[33]
266	Metabolic: HI	Metabolic	N	505	2,325	0.0183	[33]
267	Metabolic: HP	Metabolic	N	365	1,703	0.0256	[33]
268	Metabolic: MB	Metabolic	N	418	1,850	0.0212	[33]
269	Metabolic: MG	Metabolic	N	199	783	0.0397	[33]
270	Metabolic: MJ	Metabolic	N	422	1,874	0.0211	[33]
271	Metabolic: ML	Metabolic	N	414	1,862	0.0218	[33]
272	Metabolic: MP	Metabolic	N	171	685	0.0471	[33]
273	Metabolic: MT	Metabolic	N	577	2,653	0.0160	[33]
274	Metabolic: NG	Metabolic	N	394	1,824	0.0236	[33]
275	Metabolic: NM	Metabolic	N	369	1,708	0.0252	[33]
276	Metabolic: OS	Metabolic	N	285	1,168	0.0289	[33]
277	Metabolic: PA	Metabolic	N	720	3,429	0.0132	[33]
278	Metabolic: PF	Metabolic	N	310	1,379	0.0288	[33]
279	Metabolic: PG	Metabolic	N	412	1,772	0.0209	[33]
280	Metabolic: PH	Metabolic	N	318	1,394	0.0277	[33]
281	Metabolic: PN	Metabolic	N	405	1,829	0.0224	[33]
282	Metabolic: RC	Metabolic	N	663	3,111	0.0142	[33]
283	Metabolic: RP	Metabolic	N	203	775	0.0378	[33]
284	Metabolic: SC	Metabolic	N	552	2,595	0.0171	[33]
285	Metabolic: ST	Metabolic	N	391	1,756	0.0230	[33]
286	Metabolic: TH	Metabolic	N	427	1,955	0.0215	[33]
287	Metabolic: TM	Metabolic	N	328	1,452	0.0271	[33]
288	Metabolic: TP	Metabolic	N	194	788	0.0421	[33]
289	Metabolic: TY	Metabolic	N	803	3,863	0.0120	[33]
290	Metabolic: YP	Metabolic	N	552	2,471	0.0162	[33]
291	US political books co-purchase*	Other	N	105	441	0.0808	[34]
292	Power grid	Other	N	4,941	6,594	0.0005	[2]
293	Slovenian magazine co-purchase	Other	Y	124	5,972	0.7831	[35]
294	Transcription: <i>E. coli</i>	Other	N	328	456	0.0085	[36]
295	Transcription: Yeast	Other	N	662	1,062	0.0049	[37]
296	US airlines	Other	Y	324	2,081	0.0398	[16, 17]
297	2008 NCAA football schedule*	Other	Y	121	764	0.1052	[38]
298	Internet: autonomous systems	Other	N	22,963	48,436	0.0002	[39]
299	Garfield: scientometrics citations	Other	Y	2,678	10,368	0.0029	[40]
300	Garfield: Small and Griffith citations	Other	Y	1,024	4,916	0.0094	[40]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
301	Garfield: small-world citations	Other	N	233	994	0.0368	[40]
302	Electronic circuit (s208)*	Other	N	122	189	0.0256	[32]
303	Electronic circuit (s420)	Other	N	252	399	0.0126	[32]
304	Electronic circuit (s838)	Other	N	512	819	0.0063	[32]
305	Protein: serine protease inhibitor (1EAW)*	Other	N	53	123	0.0893	[32]
306	Protein: immunoglobulin (1A4J)*	Other	N	95	213	0.0477	[32]
307	Protein: oxidoreductase (1AOR)*	Other	N	97	212	0.0455	[32]
308	AIDS blogs*	Other	N	146	180	0.0170	[41]
309	Political blogs	Other	Y	1,222	16,714	0.0224	[42]
310	WWW (Stanford)	Other	N	8,929	26,320	0.0007	[43]
311	Trade product proximity	Other	Y	775	283,094	0.9439	[44]
312	World trade in metal (1994): Net	Other	Y	80	875	0.2769	[16, 45]
313	World trade in metal (1994): Total	Other	Y	80	875	0.2769	[16, 45]
314	Bill cosponsorship: US House 96	Political: cosponsorship	Y	438	95,529	0.9982	[46, 47]
315	Bill cosponsorship: US House 97	Political: cosponsorship	Y	435	94,374	0.9998	[46, 47]
316	Bill cosponsorship: US House 98	Political: cosponsorship	Y	437	95,256	0.9999	[46, 47]
317	Bill cosponsorship: US House 99	Political: cosponsorship	Y	437	94,999	0.9972	[46, 47]
318	Bill cosponsorship: US House 100	Political: cosponsorship	Y	439	96,125	0.9998	[46, 47]
319	Bill cosponsorship: US House 101	Political: cosponsorship	Y	437	95,263	1.0000	[46, 47]
320	Bill cosponsorship: US House 102	Political: cosponsorship	Y	437	95,051	0.9977	[46, 47]
321	Bill cosponsorship: US House 103	Political: cosponsorship	Y	437	95,028	0.9975	[46, 47]
322	Bill cosponsorship: US House 104	Political: cosponsorship	Y	439	95,925	0.9978	[46, 47]
323	Bill cosponsorship: US House 105	Political: cosponsorship	Y	442	97,373	0.9991	[46, 47]
324	Bill cosponsorship: US House 106	Political: cosponsorship	Y	436	94,820	0.9999	[46, 47]
325	Bill cosponsorship: US House 107	Political: cosponsorship	Y	442	97,233	0.9977	[46, 47]
326	Bill cosponsorship: US House 108	Political: cosponsorship	Y	439	96,104	0.9996	[46, 47]
327	Bill cosponsorship: US Senate 96	Political: cosponsorship	Y	101	5,050	1.0000	[46, 47]
328	Bill cosponsorship: US Senate 97	Political: cosponsorship	Y	101	5,050	1.0000	[46, 47]
329	Bill cosponsorship: US Senate 98	Political: cosponsorship	Y	101	5,050	1.0000	[46, 47]
330	Bill cosponsorship: US Senate 99	Political: cosponsorship	Y	101	5,049	0.9998	[46, 47]
331	Bill cosponsorship: US Senate 100	Political: cosponsorship	Y	101	5,050	1.0000	[46, 47]
332	Bill cosponsorship: US Senate 101	Political: cosponsorship	Y	100	4,950	1.0000	[46, 47]
333	Bill cosponsorship: US Senate 102	Political: cosponsorship	Y	102	5,142	0.9983	[46, 47]
334	Bill cosponsorship: US Senate 103	Political: cosponsorship	Y	101	5,050	1.0000	[46, 47]
335	Bill cosponsorship: US Senate 104	Political: cosponsorship	Y	102	5,151	1.0000	[46, 47]
336	Bill cosponsorship: US Senate 105	Political: cosponsorship	Y	100	4,950	1.0000	[46, 47]
337	Bill cosponsorship: US Senate 106	Political: cosponsorship	Y	102	5,151	1.0000	[46, 47]
338	Bill cosponsorship: US Senate 107	Political: cosponsorship	Y	101	5,049	0.9998	[46, 47]
339	Bill cosponsorship: US Senate 108	Political: cosponsorship	Y	100	4,950	1.0000	[46, 47]
340	Committees: US House 101, comms.	Political: committee	N	159	3,610	0.2874	[48, 49]
341	Committees: US House 102, comms.	Political: committee	N	163	4,093	0.3100	[48, 49]
342	Committees: US House 103, comms.	Political: committee	N	141	2,983	0.3022	[48, 49]
343	Committees: US House 104, comms.	Political: committee	N	106	1,839	0.3305	[48, 49]
344	Committees: US House 105, comms.	Political: committee	N	108	1,997	0.3456	[48, 49]
345	Committees: US House 106, comms.	Political: committee	N	107	2,031	0.3581	[48, 49]
346	Committees: US House 107, comms.	Political: committee	N	113	2,429	0.3838	[48, 49]
347	Committees: US House 108, comms.	Political: committee	N	118	2,905	0.4208	[48, 49]
348	Committees: US House 101, Reps.	Political: committee	N	434	18,714	0.1992	[48, 49]
349	Committees: US House 102, Reps.	Political: committee	N	436	20,134	0.2123	[48, 49]
350	Committees: US House 103, Reps.	Political: committee	N	437	18,212	0.1912	[48, 49]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
351	Committees: US House 104, Reps.	Political: committee	N	432	17,130	0.1840	[48, 49]
352	Committees: US House 105, Reps.	Political: committee	N	435	18,297	0.1938	[48, 49]
353	Committees: US House 106, Reps.	Political: committee	N	435	18,832	0.1995	[48, 49]
354	Committees: US House 107, Reps.	Political: committee	N	434	19,824	0.2110	[48, 49]
355	Committees: US House 108, Reps.	Political: committee	N	437	21,214	0.2227	[48, 49]
356	Roll call: US House 1	Political: voting	Y	66	2,122	0.9893	[50–52]
357	Roll call: US House 2	Political: voting	Y	71	2,428	0.9771	[50–52]
358	Roll call: US House 3	Political: voting	Y	108	5,669	0.9811	[50–52]
359	Roll call: US House 4	Political: voting	Y	114	6,342	0.9846	[50–52]
360	Roll call: US House 5	Political: voting	Y	117	6,600	0.9726	[50–52]
361	Roll call: US House 6	Political: voting	Y	113	6,222	0.9832	[50–52]
362	Roll call: US House 7	Political: voting	Y	110	5,921	0.9877	[50–52]
363	Roll call: US House 8	Political: voting	Y	149	10,888	0.9875	[50–52]
364	Roll call: US House 9	Political: voting	Y	147	10,582	0.9861	[50–52]
365	Roll call: US House 10	Political: voting	Y	149	10,857	0.9847	[50–52]
366	Roll call: US House 11	Political: voting	Y	153	11,482	0.9874	[50–52]
367	Roll call: US House 12	Political: voting	Y	146	10,535	0.9953	[50–52]
368	Roll call: US House 13	Political: voting	Y	195	18,723	0.9898	[50–52]
369	Roll call: US House 14	Political: voting	Y	195	18,540	0.9802	[50–52]
370	Roll call: US House 15	Political: voting	Y	195	18,666	0.9868	[50–52]
371	Roll call: US House 16	Political: voting	Y	197	19,118	0.9903	[50–52]
372	Roll call: US House 17	Political: voting	Y	199	19,429	0.9862	[50–52]
373	Roll call: US House 18	Political: voting	Y	221	23,812	0.9795	[50–52]
374	Roll call: US House 19	Political: voting	Y	220	23,993	0.9960	[50–52]
375	Roll call: US House 20	Political: voting	Y	219	23,666	0.9914	[50–52]
376	Roll call: US House 21	Political: voting	Y	220	23,985	0.9956	[50–52]
377	Roll call: US House 22	Political: voting	Y	217	23,404	0.9986	[50–52]
378	Roll call: US House 23	Political: voting	Y	257	32,502	0.9880	[50–52]
379	Roll call: US House 24	Political: voting	Y	255	32,062	0.9900	[50–52]
380	Roll call: US House 25	Political: voting	Y	256	32,366	0.9916	[50–52]
381	Roll call: US House 26	Political: voting	Y	255	32,067	0.9902	[50–52]
382	Roll call: US House 27	Political: voting	Y	257	32,743	0.9953	[50–52]
383	Roll call: US House 28	Political: voting	Y	234	26,788	0.9826	[50–52]
384	Roll call: US House 29	Political: voting	Y	236	27,562	0.9939	[50–52]
385	Roll call: US House 30	Political: voting	Y	236	27,669	0.9978	[50–52]
386	Roll call: US House 31	Political: voting	Y	241	28,804	0.9960	[50–52]
387	Roll call: US House 32	Political: voting	Y	239	28,318	0.9957	[50–52]
388	Roll call: US House 33	Political: voting	Y	240	28,570	0.9962	[50–52]
389	Roll call: US House 34	Political: voting	Y	236	27,545	0.9933	[50–52]
390	Roll call: US House 35	Political: voting	Y	245	29,630	0.9913	[50–52]
391	Roll call: US House 36	Political: voting	Y	243	29,312	0.9969	[50–52]
392	Roll call: US House 37	Political: voting	Y	197	18,735	0.9704	[50–52]
393	Roll call: US House 38	Political: voting	Y	187	17,326	0.9963	[50–52]
394	Roll call: US House 39	Political: voting	Y	199	19,593	0.9945	[50–52]
395	Roll call: US House 40	Political: voting	Y	233	26,605	0.9843	[50–52]
396	Roll call: US House 41	Political: voting	Y	256	32,109	0.9837	[50–52]
397	Roll call: US House 42	Political: voting	Y	253	31,626	0.9921	[50–52]
398	Roll call: US House 43	Political: voting	Y	302	45,151	0.9934	[50–52]
399	Roll call: US House 44	Political: voting	Y	308	46,723	0.9883	[50–52]
400	Roll call: US House 45	Political: voting	Y	302	45,315	0.9970	[50–52]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
401	Roll call: US House 46	Political: voting	Y	301	44,987	0.9964	[50–52]
402	Roll call: US House 47	Political: voting	Y	306	46,214	0.9903	[50–52]
403	Roll call: US House 48	Political: voting	Y	338	56,484	0.9918	[50–52]
404	Roll call: US House 49	Political: voting	Y	330	54,160	0.9977	[50–52]
405	Roll call: US House 50	Political: voting	Y	326	52,907	0.9987	[50–52]
406	Roll call: US House 51	Political: voting	Y	347	59,303	0.9879	[50–52]
407	Roll call: US House 52	Political: voting	Y	340	57,285	0.9940	[50–52]
408	Roll call: US House 53	Political: voting	Y	376	69,943	0.9921	[50–52]
409	Roll call: US House 54	Political: voting	Y	368	67,085	0.9934	[50–52]
410	Roll call: US House 55	Political: voting	Y	371	68,270	0.9947	[50–52]
411	Roll call: US House 56	Political: voting	Y	369	67,059	0.9877	[50–52]
412	Roll call: US House 57	Political: voting	Y	371	67,383	0.9818	[50–52]
413	Roll call: US House 58	Political: voting	Y	397	75,891	0.9655	[50–52]
414	Roll call: US House 59	Political: voting	Y	397	76,299	0.9707	[50–52]
415	Roll call: US House 60	Political: voting	Y	398	77,921	0.9863	[50–52]
416	Roll call: US House 61	Political: voting	Y	402	80,174	0.9947	[50–52]
417	Roll call: US House 62	Political: voting	Y	408	82,442	0.9929	[50–52]
418	Roll call: US House 63	Political: voting	Y	452	101,498	0.9958	[50–52]
419	Roll call: US House 64	Political: voting	Y	441	96,780	0.9975	[50–52]
420	Roll call: US House 65	Political: voting	Y	454	102,108	0.9930	[50–52]
421	Roll call: US House 66	Political: voting	Y	453	101,199	0.9885	[50–52]
422	Roll call: US House 67	Political: voting	Y	452	101,482	0.9956	[50–52]
423	Roll call: US House 68	Political: voting	Y	442	96,885	0.9941	[50–52]
424	Roll call: US House 69	Political: voting	Y	437	95,226	0.9996	[50–52]
425	Roll call: US House 70	Political: voting	Y	443	97,497	0.9959	[50–52]
426	Roll call: US House 71	Political: voting	Y	455	102,502	0.9924	[50–52]
427	Roll call: US House 72	Political: voting	Y	447	99,028	0.9934	[50–52]
428	Roll call: US House 73	Political: voting	Y	445	98,647	0.9986	[50–52]
429	Roll call: US House 74	Political: voting	Y	440	96,170	0.9958	[50–52]
430	Roll call: US House 75	Political: voting	Y	445	98,474	0.9968	[50–52]
431	Roll call: US House 76	Political: voting	Y	456	102,495	0.9880	[50–52]
432	Roll call: US House 77	Political: voting	Y	450	99,956	0.9894	[50–52]
433	Roll call: US House 78	Political: voting	Y	450	100,513	0.9949	[50–52]
434	Roll call: US House 79	Political: voting	Y	448	99,246	0.9912	[50–52]
435	Roll call: US House 80	Political: voting	Y	448	99,902	0.9977	[50–52]
436	Roll call: US House 81	Political: voting	Y	444	98,054	0.9970	[50–52]
437	Roll call: US House 82	Political: voting	Y	447	99,281	0.9960	[50–52]
438	Roll call: US House 83	Political: voting	Y	440	96,506	0.9992	[50–52]
439	Roll call: US House 84	Political: voting	Y	437	95,253	0.9999	[50–52]
440	Roll call: US House 85	Political: voting	Y	444	97,955	0.9960	[50–52]
441	Roll call: US House 86	Political: voting	Y	443	97,377	0.9946	[50–52]
442	Roll call: US House 87	Political: voting	Y	449	99,774	0.9920	[50–52]
443	Roll call: US House 88	Political: voting	Y	443	97,842	0.9994	[50–52]
444	Roll call: US House 89	Political: voting	Y	442	97,139	0.9967	[50–52]
445	Roll call: US House 90	Political: voting	Y	437	95,251	0.9998	[50–52]
446	Roll call: US House 91	Political: voting	Y	448	99,815	0.9969	[50–52]
447	Roll call: US House 92	Political: voting	Y	443	97,579	0.9967	[50–52]
448	Roll call: US House 93	Political: voting	Y	443	97,848	0.9994	[50–52]
449	Roll call: US House 94	Political: voting	Y	441	96,837	0.9981	[50–52]
450	Roll call: US House 95	Political: voting	Y	441	96,493	0.9946	[50–52]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
451	Roll call: US House 96	Political: voting	Y	440	96,379	0.9979	[50–52]
452	Roll call: US House 97	Political: voting	Y	442	96,761	0.9928	[50–52]
453	Roll call: US House 98	Political: voting	Y	439	95,922	0.9977	[50–52]
454	Roll call: US House 99	Political: voting	Y	439	95,875	0.9972	[50–52]
455	Roll call: US House 100	Political: voting	Y	440	96,544	0.9996	[50–52]
456	Roll call: US House 101	Political: voting	Y	440	96,505	0.9992	[50–52]
457	Roll call: US House 102	Political: voting	Y	441	96,811	0.9978	[50–52]
458	Roll call: US House 103	Political: voting	Y	441	96,348	0.9931	[50–52]
459	Roll call: US House 104	Political: voting	Y	445	98,720	0.9993	[50–52]
460	Roll call: US House 105	Political: voting	Y	443	97,841	0.9994	[50–52]
461	Roll call: US House 106	Political: voting	Y	440	96,557	0.9998	[50–52]
462	Roll call: US House 107	Political: voting	Y	443	97,816	0.9991	[50–52]
463	Roll call: US House 108	Political: voting	Y	440	96,561	0.9998	[50–52]
464	Roll call: US House 109	Political: voting	Y	440	96,549	0.9997	[50–52]
465	Roll call: US House 110	Political: voting	Y	448	99,603	0.9948	[50–52]
466	Roll call: US Senate 1	Political: voting	Y	29	393	0.9680	[50–52]
467	Roll call: US Senate 2	Political: voting	Y	31	449	0.9656	[50–52]
468	Roll call: US Senate 3	Political: voting	Y	32	472	0.9516	[50–52]
469	Roll call: US Senate 4	Political: voting	Y	43	760	0.8416	[50–52]
470	Roll call: US Senate 5	Political: voting	Y	44	808	0.8541	[50–52]
471	Roll call: US Senate 6	Political: voting	Y	37	644	0.9670	[50–52]
472	Roll call: US Senate 7	Political: voting	Y	35	537	0.9025	[50–52]
473	Roll call: US Senate 8	Political: voting	Y	44	864	0.9133	[50–52]
474	Roll call: US Senate 9	Political: voting	Y	37	645	0.9685	[50–52]
475	Roll call: US Senate 10	Political: voting	Y	37	660	0.9910	[50–52]
476	Roll call: US Senate 11	Political: voting	Y	44	855	0.9038	[50–52]
477	Roll call: US Senate 12	Political: voting	Y	37	663	0.9955	[50–52]
478	Roll call: US Senate 13	Political: voting	Y	46	947	0.9150	[50–52]
479	Roll call: US Senate 14	Political: voting	Y	44	898	0.9493	[50–52]
480	Roll call: US Senate 15	Political: voting	Y	46	977	0.9440	[50–52]
481	Roll call: US Senate 16	Political: voting	Y	51	1,249	0.9796	[50–52]
482	Roll call: US Senate 17	Political: voting	Y	52	1,294	0.9759	[50–52]
483	Roll call: US Senate 18	Political: voting	Y	52	1,304	0.9834	[50–52]
484	Roll call: US Senate 19	Political: voting	Y	59	1,589	0.9287	[50–52]
485	Roll call: US Senate 20	Political: voting	Y	53	1,343	0.9746	[50–52]
486	Roll call: US Senate 21	Political: voting	Y	54	1,339	0.9357	[50–52]
487	Roll call: US Senate 22	Political: voting	Y	53	1,348	0.9782	[50–52]
488	Roll call: US Senate 23	Political: voting	Y	54	1,378	0.9630	[50–52]
489	Roll call: US Senate 24	Political: voting	Y	61	1,732	0.9464	[50–52]
490	Roll call: US Senate 25	Political: voting	Y	58	1,627	0.9843	[50–52]
491	Roll call: US Senate 26	Political: voting	Y	60	1,689	0.9542	[50–52]
492	Roll call: US Senate 27	Political: voting	Y	59	1,662	0.9714	[50–52]
493	Roll call: US Senate 28	Political: voting	Y	57	1,575	0.9868	[50–52]
494	Roll call: US Senate 29	Political: voting	Y	63	1,895	0.9703	[50–52]
495	Roll call: US Senate 30	Political: voting	Y	72	2,320	0.9077	[50–52]
496	Roll call: US Senate 31	Political: voting	Y	70	2,341	0.9694	[50–52]
497	Roll call: US Senate 32	Political: voting	Y	73	2,511	0.9555	[50–52]
498	Roll call: US Senate 33	Political: voting	Y	70	2,308	0.9557	[50–52]
499	Roll call: US Senate 34	Political: voting	Y	64	2,002	0.9931	[50–52]
500	Roll call: US Senate 35	Political: voting	Y	73	2,542	0.9673	[50–52]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
501	Roll call: US Senate 36	Political: voting	Y	70	2,370	0.9814	[50–52]
502	Roll call: US Senate 37	Political: voting	Y	70	2,051	0.8493	[50–52]
503	Roll call: US Senate 38	Political: voting	Y	54	1,402	0.9797	[50–52]
504	Roll call: US Senate 39	Political: voting	Y	59	1,610	0.9410	[50–52]
505	Roll call: US Senate 40	Political: voting	Y	69	2,274	0.9693	[50–52]
506	Roll call: US Senate 41	Political: voting	Y	80	3,084	0.9759	[50–52]
507	Roll call: US Senate 42	Political: voting	Y	75	2,773	0.9993	[50–52]
508	Roll call: US Senate 43	Political: voting	Y	79	3,041	0.9870	[50–52]
509	Roll call: US Senate 44	Political: voting	Y	82	3,261	0.9819	[50–52]
510	Roll call: US Senate 45	Political: voting	Y	82	3,265	0.9831	[50–52]
511	Roll call: US Senate 46	Political: voting	Y	81	3,219	0.9935	[50–52]
512	Roll call: US Senate 47	Political: voting	Y	83	3,362	0.9880	[50–52]
513	Roll call: US Senate 48	Political: voting	Y	78	2,998	0.9983	[50–52]
514	Roll call: US Senate 49	Political: voting	Y	81	3,210	0.9907	[50–52]
515	Roll call: US Senate 50	Political: voting	Y	76	2,850	1.0000	[50–52]
516	Roll call: US Senate 51	Political: voting	Y	91	3,998	0.9763	[50–52]
517	Roll call: US Senate 52	Political: voting	Y	93	4,249	0.9932	[50–52]
518	Roll call: US Senate 53	Political: voting	Y	95	4,413	0.9884	[50–52]
519	Roll call: US Senate 54	Political: voting	Y	90	4,000	0.9988	[50–52]
520	Roll call: US Senate 55	Political: voting	Y	96	4,445	0.9748	[50–52]
521	Roll call: US Senate 56	Political: voting	Y	93	4,201	0.9820	[50–52]
522	Roll call: US Senate 57	Political: voting	Y	90	3,939	0.9835	[50–52]
523	Roll call: US Senate 58	Political: voting	Y	93	4,174	0.9757	[50–52]
524	Roll call: US Senate 59	Political: voting	Y	93	4,251	0.9937	[50–52]
525	Roll call: US Senate 60	Political: voting	Y	95	4,382	0.9814	[50–52]
526	Roll call: US Senate 61	Political: voting	Y	102	5,033	0.9771	[50–52]
527	Roll call: US Senate 62	Political: voting	Y	109	5,719	0.9716	[50–52]
528	Roll call: US Senate 63	Political: voting	Y	101	5,029	0.9958	[50–52]
529	Roll call: US Senate 64	Political: voting	Y	100	4,931	0.9962	[50–52]
530	Roll call: US Senate 65	Political: voting	Y	111	5,899	0.9663	[50–52]
531	Roll call: US Senate 66	Political: voting	Y	101	5,005	0.9911	[50–52]
532	Roll call: US Senate 67	Political: voting	Y	105	5,413	0.9914	[50–52]
533	Roll call: US Senate 68	Political: voting	Y	102	5,081	0.9864	[50–52]
534	Roll call: US Senate 69	Political: voting	Y	105	5,353	0.9804	[50–52]
535	Roll call: US Senate 70	Political: voting	Y	102	5,082	0.9866	[50–52]
536	Roll call: US Senate 71	Political: voting	Y	109	5,779	0.9818	[50–52]
537	Roll call: US Senate 72	Political: voting	Y	103	5,220	0.9937	[50–52]
538	Roll call: US Senate 73	Political: voting	Y	100	4,879	0.9857	[50–52]
539	Roll call: US Senate 74	Political: voting	Y	100	4,933	0.9966	[50–52]
540	Roll call: US Senate 75	Political: voting	Y	102	5,126	0.9951	[50–52]
541	Roll call: US Senate 76	Political: voting	Y	104	5,106	0.9533	[50–52]
542	Roll call: US Senate 77	Political: voting	Y	108	5,575	0.9649	[50–52]
543	Roll call: US Senate 78	Political: voting	Y	104	5,304	0.9903	[50–52]
544	Roll call: US Senate 79	Political: voting	Y	107	5,466	0.9639	[50–52]
545	Roll call: US Senate 80	Political: voting	Y	97	4,655	0.9998	[50–52]
546	Roll call: US Senate 81	Political: voting	Y	108	5,646	0.9772	[50–52]
547	Roll call: US Senate 82	Political: voting	Y	98	4,748	0.9989	[50–52]
548	Roll call: US Senate 83	Political: voting	Y	110	5,724	0.9548	[50–52]
549	Roll call: US Senate 84	Political: voting	Y	99	4,845	0.9988	[50–52]
550	Roll call: US Senate 85	Political: voting	Y	101	5,014	0.9929	[50–52]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
551	Roll call: US Senate 86	Political: voting	Y	103	5,246	0.9987	[50–52]
552	Roll call: US Senate 87	Political: voting	Y	105	5,444	0.9971	[50–52]
553	Roll call: US Senate 88	Political: voting	Y	103	5,249	0.9992	[50–52]
554	Roll call: US Senate 89	Political: voting	Y	103	5,247	0.9989	[50–52]
555	Roll call: US Senate 90	Political: voting	Y	101	5,048	0.9996	[50–52]
556	Roll call: US Senate 91	Political: voting	Y	102	5,148	0.9994	[50–52]
557	Roll call: US Senate 92	Political: voting	Y	102	5,147	0.9992	[50–52]
558	Roll call: US Senate 93	Political: voting	Y	103	5,246	0.9987	[50–52]
559	Roll call: US Senate 94	Political: voting	Y	101	5,049	0.9998	[50–52]
560	Roll call: US Senate 95	Political: voting	Y	104	5,345	0.9979	[50–52]
561	Roll call: US Senate 96	Political: voting	Y	101	5,049	0.9998	[50–52]
562	Roll call: US Senate 97	Political: voting	Y	101	5,049	0.9998	[50–52]
563	Roll call: US Senate 98	Political: voting	Y	101	5,049	0.9998	[50–52]
564	Roll call: US Senate 99	Political: voting	Y	101	5,049	0.9998	[50–52]
565	Roll call: US Senate 100	Political: voting	Y	101	5,049	0.9998	[50–52]
566	Roll call: US Senate 101	Political: voting	Y	100	4,950	1.0000	[50–52]
567	Roll call: US Senate 102	Political: voting	Y	102	5,148	0.9994	[50–52]
568	Roll call: US Senate 103	Political: voting	Y	102	5,080	0.9862	[50–52]
569	Roll call: US Senate 104	Political: voting	Y	103	5,247	0.9989	[50–52]
570	Roll call: US Senate 105	Political: voting	Y	100	4,950	1.0000	[50–52]
571	Roll call: US Senate 106	Political: voting	Y	102	5,148	0.9994	[50–52]
571	Roll call: US Senate 107	Political: voting	Y	102	5,148	0.9994	[50–52]
573	Roll call: US Senate 108	Political: voting	Y	100	4,950	1.0000	[50–52]
574	Roll call: US Senate 109	Political: voting	Y	101	5,049	0.9998	[50–52]
575	Roll call: US Senate 110	Political: voting	Y	102	5,147	0.9992	[50–52]
576	UK House of Commons voting: 1992-1997	Political: voting	Y	668	220,761	0.9909	[53]
577	UK House of Commons voting: 1997-2001	Political: voting	Y	671	223,092	0.9925	[53]
578	UK House of Commons voting: 2001-2005	Political: voting	Y	657	215,246	0.9988	[53]
579	UN resolutions 1	Political: voting	Y	54	1,431	1.0000	[54]
580	UN resolutions 2	Political: voting	Y	57	1,594	0.9987	[54]
581	UN resolutions 3	Political: voting	Y	59	1,711	1.0000	[54]
582	UN resolutions 4	Political: voting	Y	59	1,711	1.0000	[54]
583	UN resolutions 5	Political: voting	Y	60	1,770	1.0000	[54]
584	UN resolutions 6	Political: voting	Y	60	1,768	0.9989	[54]
585	UN resolutions 7	Political: voting	Y	60	1,770	1.0000	[54]
586	UN resolutions 8	Political: voting	Y	60	1,770	1.0000	[54]
587	UN resolutions 9	Political: voting	Y	60	1,770	1.0000	[54]
588	UN resolutions 10	Political: voting	Y	65	2,037	0.9793	[54]
589	UN resolutions 11	Political: voting	Y	81	3,239	0.9997	[54]
590	UN resolutions 12	Political: voting	Y	82	3,317	0.9988	[54]
591	UN resolutions 13	Political: voting	Y	82	3,294	0.9919	[54]
592	UN resolutions 14	Political: voting	Y	82	3,321	1.0000	[54]
593	UN resolutions 15	Political: voting	Y	99	4,851	1.0000	[54]
594	UN resolutions 16	Political: voting	Y	104	5,356	1.0000	[54]
595	UN resolutions 17	Political: voting	Y	110	5,995	1.0000	[54]
596	UN resolutions 18	Political: voting	Y	113	6,246	0.9870	[54]
597	UN resolutions 20	Political: voting	Y	117	6,672	0.9832	[54]
598	UN resolutions 21	Political: voting	Y	122	7,333	0.9935	[54]
599	UN resolutions 22	Political: voting	Y	124	7,616	0.9987	[54]
600	UN resolutions 23	Political: voting	Y	126	7,855	0.9975	[54]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
601	UN resolutions 24	Political: voting	Y	126	7,851	0.9970	[54]
602	UN resolutions 25	Political: voting	Y	126	7,868	0.9991	[54]
603	UN resolutions 26	Political: voting	Y	132	8,641	0.9994	[54]
604	UN resolutions 27	Political: voting	Y	132	8,646	1.0000	[54]
605	UN resolutions 28	Political: voting	Y	134	8,905	0.9993	[54]
606	UN resolutions 29	Political: voting	Y	137	9,202	0.9878	[54]
607	UN resolutions 30	Political: voting	Y	143	10,117	0.9965	[54]
608	UN resolutions 31	Political: voting	Y	144	10,291	0.9995	[54]
609	UN resolutions 32	Political: voting	Y	146	10,585	1.0000	[54]
610	UN resolutions 33	Political: voting	Y	148	10,878	1.0000	[54]
611	UN resolutions 34	Political: voting	Y	150	11,173	0.9998	[54]
612	UN resolutions 35	Political: voting	Y	151	11,287	0.9966	[54]
613	UN resolutions 36	Political: voting	Y	155	11,935	1.0000	[54]
614	UN resolutions 37	Political: voting	Y	156	12,090	1.0000	[54]
615	UN resolutions 38	Political: voting	Y	157	12,243	0.9998	[54]
616	UN resolutions 39	Political: voting	Y	158	12,403	1.0000	[54]
617	UN resolutions 40	Political: voting	Y	158	12,403	1.0000	[54]
618	UN resolutions 41	Political: voting	Y	158	12,403	1.0000	[54]
619	UN resolutions 42	Political: voting	Y	158	12,402	0.9999	[54]
620	UN resolutions 43	Political: voting	Y	158	12,403	1.0000	[54]
621	UN resolutions 44	Political: voting	Y	158	12,403	1.0000	[54]
622	UN resolutions 45	Political: voting	Y	154	11,781	1.0000	[54]
623	UN resolutions 46	Political: voting	Y	168	13,872	0.9889	[54]
624	UN resolutions 47	Political: voting	Y	174	14,944	0.9929	[54]
625	UN resolutions 48	Political: voting	Y	178	15,606	0.9907	[54]
626	UN resolutions 49	Political: voting	Y	174	14,913	0.9908	[54]
627	UN resolutions 50	Political: voting	Y	179	15,826	0.9934	[54]
628	UN resolutions 51	Political: voting	Y	180	16,096	0.9991	[54]
629	UN resolutions 52	Political: voting	Y	176	15,349	0.9967	[54]
630	UN resolutions 53	Political: voting	Y	177	15,500	0.9951	[54]
631	UN resolutions 54	Political: voting	Y	174	14,970	0.9946	[54]
632	UN resolutions 55	Political: voting	Y	182	16,333	0.9916	[54]
633	UN resolutions 56	Political: voting	Y	179	15,812	0.9925	[54]
634	UN resolutions 57	Political: voting	Y	187	17,373	0.9990	[54]
635	UN resolutions 58	Political: voting	Y	189	17,735	0.9983	[54]
636	UN resolutions 59	Political: voting	Y	191	18,140	0.9997	[54]
637	UN resolutions 60	Political: voting	Y	191	18,110	0.9981	[54]
638	UN resolutions 61	Political: voting	Y	192	18,331	0.9997	[54]
639	UN resolutions 62	Political: voting	Y	192	18,331	0.9997	[54]
640	UN resolutions 63	Political: voting	Y	192	18,328	0.9996	[54]
641	Biogrid: <i>A. thaliana</i>	Protein interaction	N	406	625	0.0076	[55]
642	Biogrid: <i>C. elegans</i>	Protein interaction	N	3,353	6,449	0.0011	[55]
643	Biogrid: <i>D. melanogaster</i>	Protein interaction	N	7,174	24,897	0.0010	[55]
644	Biogrid: <i>H. sapien</i>	Protein interaction	N	8,205	25,699	0.0008	[55]
645	Biogrid: <i>M. musculus</i>	Protein interaction	N	710	1,003	0.0040	[55]
646	Biogrid: <i>R. norvegicus</i> *	Protein interaction	N	121	135	0.0186	[55]
647	Biogrid: <i>S. cerevisiae</i>	Protein interaction	N	1,753	4,811	0.0031	[55]
648	Biogrid: <i>S. pombe</i>	Protein interaction	N	1,477	11,404	0.0105	[55]
649	DIP: <i>H. pylori</i>	Protein interaction	N	686	1,351	0.0058	[56, 57]
650	DIP: <i>H. sapien</i>	Protein interaction	N	639	982	0.0048	[56, 57]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
651	DIP: <i>M. musculus</i>	Protein interaction	N	50	55	0.0449	[56, 57]
651	DIP: <i>C. elegans</i>	Protein interaction	N	2,386	3,825	0.0013	[56, 57]
653	Human: Ccsb	Protein interaction	N	1,307	2,483	0.0029	[58]
654	Human: Ophid	Protein interaction	N	5,464	23,238	0.0016	[59, 60]
655	STRING: <i>C. elegans</i>	Protein interaction	N	1,762	95,227	0.0614	[61]
656	STRING: <i>S. cerevisiae</i>	Protein interaction	N	534	57,672	0.4053	[61]
657	Yeast: Oxford Statistics	Protein interaction	N	2,224	6,609	0.0027	[62]
658	Yeast: DIP	Protein interaction	N	4,906	17,218	0.0014	[56, 57, 62]
659	Yeast: DIPC	Protein interaction	N	2,587	6,094	0.0018	[56, 57, 62]
660	Yeast: FHC	Protein interaction	N	2,233	5,750	0.0023	[62, 63]
661	Yeast: FYI	Protein interaction	N	778	1,798	0.0059	[62, 64]
662	Yeast: PCA	Protein interaction	N	889	2,407	0.0061	[62, 65]
663	Corporate directors in Scotland (1904-1905)*	Social	Y	131	676	0.0794	[16, 66]
664	Corporate ownership (EVA)	Social	N	4,475	4,652	0.0005	[67]
665	Dolphins*	Social	N	62	159	0.0841	[68]
666	Family planning in Korea	Social	N	33	68	0.1288	[69]
667	Unionization in a high-tech firm*	Social	N	33	91	0.1723	[70]
668	Communication within a sawmill on strike*	Social	N	36	62	0.0984	[71]
669	Leadership course	Social	N	32	80	0.1613	[32]
670	Les Miserables*	Social	Y	77	254	0.0868	[31]
671	Marvel comics	Social	Y	6,449	168,211	0.0081	[72]
672	Mexican political elite	Social	N	35	117	0.1966	[73]
673	Pretty-good-privacy (PGP) algorithm users	Social	N	10,680	24,316	0.0004	[74]
674	Prisoners	Social	N	67	142	0.0642	[32]
675	Bernard and Killworth fraternity: observed	Social	Y	58	967	0.5850	[75–77]
676	Bernard and Killworth fraternity: recalled	Social	Y	58	1,653	1.0000	[75–77]
677	Bernard and Killworth HAM radio: observed	Social	Y	41	153	0.1866	[78–80]
678	Bernard and Killworth HAM radio: recalled	Social	Y	44	442	0.4672	[78–80]
679	Bernard and Killworth office: observed	Social	Y	40	238	0.3051	[78–80]
680	Bernard and Killworth office: recalled	Social	Y	40	779	0.9987	[78–80]
681	Bernard and Killworth technical: observed	Social	Y	34	175	0.3119	[78–80]
681	Bernard and Killworth technical: recalled	Social	Y	34	561	1.0000	[78–80]
683	Kapferer tailor shop: instrumental (t1)	Social	N	35	76	0.1277	[81]
684	Kapferer tailor shop: instrumental (t2)	Social	N	34	93	0.1658	[81]
685	Kapferer tailor shop: associational (t1)	Social	N	39	158	0.2132	[81]
686	Kapferer tailor shop: associational (t2)	Social	N	39	223	0.3009	[81]
687	University Rovira i Virgili (Tarragona) e-mail	Social	N	1,133	5,451	0.0085	[82]
688	Zachary karate club*	Social	N	34	78	0.1390	[83]
689	BA: (100,1)*	Synthetic	N	100	99	0.0200	[3]
690	BA: (100,2)*	Synthetic	N	100	197	0.0398	[3]
691	BA: (1000,1)	Synthetic	N	1,000	999	0.0020	[3]
692	BA: (1000,2)	Synthetic	N	1,000	1,997	0.0040	[3]
693	BA: (500,1)	Synthetic	N	500	499	0.0040	[3]
694	BA: (500,2)	Synthetic	N	500	997	0.0080	[3]
695	ER: (100,0.25)*	Synthetic	N	100	1,264	0.2554	[1]
696	ER: (100,0.5)	Synthetic	N	100	2,436	0.4921	[1]
697	ER: (100,0.75)	Synthetic	N	100	3,697	0.7469	[1]
698	ER: (500,0.25)	Synthetic	N	500	31,148	0.2497	[1]
699	ER: (500,0.5)	Synthetic	N	500	62,301	0.4994	[1]
700	ER: (500,75)	Synthetic	N	500	93,780	0.7517	[1]

SM TABLE II. (Continued.)

ID	Name	Category	Weighted	N	L	f_e	References
701	Fractal: (10,2,1)	Synthetic	N	1,024	9,256	0.0177	[4]
702	Fractal: (10,2,2)	Synthetic	N	1,024	16,875	0.0322	[4]
703	Fractal: (10,2,3)	Synthetic	N	1,024	30,344	0.0579	[4]
704	Fractal: (10,2,4)	Synthetic	N	1,024	53,009	0.1012	[4]
705	Fractal: (10,2,5)	Synthetic	N	1,024	89,812	0.1715	[4]
706	Fractal: (10,2,6)	Synthetic	N	1,024	147,784	0.2822	[4]
707	Fractal: (10,2,7)	Synthetic	N	1,024	232,794	0.4445	[4]
708	Fractal: (10,2,8)	Synthetic	N	1,024	343,563	0.6559	[4]
709	H13-4 benchmark	Synthetic	N	256	2,311	0.0708	[84]
710	LFR benchmark: (1000,15,50,0.1,2,2)	Synthetic	N	1,000	7,573	0.0152	[6]
711	LFR benchmark: (1000,15,50,0.1,3,1)	Synthetic	N	1,000	7,447	0.0149	[6]
712	LFR benchmark: (1000,15,50,0.5,2,2)	Synthetic	N	1,000	7,624	0.0153	[6]
713	LFR benchmark: (1000,15,50,0.5,3,1)	Synthetic	N	1,000	7,177	0.0144	[6]
714	LFR benchmark: (1000,25,50,0.1,2,2)	Synthetic	N	1,000	12,739	0.0255	[6]
715	LFR benchmark: (1000,25,50,0.1,3,1)	Synthetic	N	1,000	12,523	0.0251	[6]
716	LFR benchmark: (1000,25,50,0.5,2,2)	Synthetic	N	1,000	12,744	0.0255	[6]
717	LFR benchmark: (1000,25,50,0.5,3,1)	Synthetic	N	1,000	12,662	0.0253	[6]
718	LF benchmark: (1000,15,50,0.1,0.1,1,2,1)	Synthetic	Y	1,000	7,680	0.0154	[8]
719	LF benchmark: (1000,15,50,0.1,0.1,1,2,2)	Synthetic	Y	1,000	7,791	0.0156	[8]
720	LF benchmark: (1000,15,50,0.5,0.1,1,2,1)	Synthetic	Y	1,000	7,657	0.0153	[8]
721	LF benchmark: (1000,15,50,0.5,0.1,2,2,2)	Synthetic	Y	1,000	7,912	0.0158	[8]
722	LF benchmark: (1000,15,50,0.5,0.5,1,2,1)	Synthetic	Y	1,000	7,693	0.0154	[8]
723	LF benchmark: (1000,15,50,0.5,0.5,1,2,2)	Synthetic	Y	1,000	7,906	0.0158	[8]
724	LF benchmark: (1000,25,50,0.1,0.1,1,2,1)	Synthetic	Y	1,000	12,660	0.0253	[8]
725	LF benchmark: (1000,25,50,0.1,0.1,2,2,2)	Synthetic	Y	1,000	12,641	0.0253	[8]
726	LF benchmark: (1000,25,50,0.5,0.1,1,2,1)	Synthetic	Y	1,000	12,771	0.0256	[8]
727	LF benchmark: (1000,25,50,0.5,0.1,2,2,2)	Synthetic	Y	1,000	12,772	0.0256	[8]
728	LF benchmark: (1000,25,50,0.5,0.5,1,2,1)	Synthetic	Y	1,000	12,962	0.0259	[8]
729	LF benchmark: (1000,25,50,0.5,0.5,2,2,2)	Synthetic	Y	1,000	12,881	0.0258	[8]
730	LF-NG benchmark	Synthetic	Y	128	1,024	0.1260	[8, 9]
731	Random fully-connected: (100)	Synthetic	Y	100	4,950	1.0000	
732	Random fully-connected: (500)	Synthetic	Y	500	124,750	1.0000	[†]
733	WS: (100,1,0.1)	Synthetic	N	100	100	0.0202	[2]
734	WS: (100,1,0.5)	Synthetic	N	73	73	0.0278	[2]
735	WS: (100,4,0.1)	Synthetic	N	100	407	0.0822	[2]
736	WS: (100,4,0.5)	Synthetic	N	100	522	0.1055	[2]
737	WS: (1000,1,0.1)	Synthetic	N	850	850	0.0024	[2]
738	WS: (1000,1,0.5)	Synthetic	N	877	877	0.0023	[2]
739	WS: (1000,4,0.1)	Synthetic	N	1,000	4,053	0.0081	[2]
740	WS: (1000,4,0.5)	Synthetic	N	1,000	5,138	0.0103	[2]
741	KOSKK:(1000,1,10,10,5 $\times 10^{-5}$,1 $\times 10^{-3}$,100)	Synthetic	Y	519	2,096	0.0156	[5]
742	KOSKK:(1000,1,10,10,5 $\times 10^{-5}$,1 $\times 10^{-3}$,1000)	Synthetic	Y	895	7,682	0.0192	[5]
743	KOSKK:(1000,1,100,10,5 $\times 10^{-5}$,1 $\times 10^{-3}$,1000)	Synthetic	Y	870	4,725	0.0125	[5]
744	KOSKK:(1000,1,100,105 $\times 10^{-5}$,1 $\times 10^{-3}$,100)	Synthetic	Y	652	2,125	0.0100	[5]
745	KOSKK:(1000,1,50,10,5 $\times 10^{-5}$,1 $\times 10^{-3}$,100)	Synthetic	Y	459	1,554	0.0148	[5]
746	KOSKK:(1000,1,50,10,5 $\times 10^{-5}$,1 $\times 10^{-3}$,1000)	Synthetic	Y	851	4,960	0.0137	[5]

- [1] P. Erdős and A. Rényi, *Publicationes Mathematicae* **6**, 290 (1959).
- [2] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [3] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [4] O. Sporns, *BioSystems* **85**, 55 (2006).
- [5] J. M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész, *Physical Review Letters* **99**, 228701 (2007).
- [6] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Physical Review E* **78**, 046110 (2008).
- [7] M. Molloy and B. Reed, *Random Structures and Algorithms* **6**, 161 (1995).
- [8] A. Lancichinetti and S. Fortunato, *Physical Review E* **80**, 016118 (2009).
- [9] M. E. J. Newman and M. Girvan, *Physical Review E* **69**, 026113 (2004).
- [10] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, and O. Sporns, *PLoS Biology* **6**, e159 (2008).
- [11] J. W. Scannell, G. A. P. C. Burns, C. C. Hilgetag, M. A. O’Neil, and M. P. Young, *Cerebral Cortex* **9**, 277 (1999).
- [12] D. J. Felleman and D. C. V. Essen, *Cerebral Cortex* **1**, 1 (1991).
- [13] M. P. Young, *Proceedings of the Royal Society B* **252**, 13 (1993).
- [14] M. E. J. Newman, *Proceedings of the National Academy of Sciences of the U.S.A.* **98**, 404 (2001).
- [15] N. H. F. Beebe, (2002), The authors collaboration network in computational geometry was produced from the BibTeX bibliography available at <http://www.math.utah.edu/~beebe/bibliographies.html>. The network data is available at <http://vlado.fmf.uni-lj.si/pub/networks/data/collab/geom.htm>.
- [16] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek* (Cambridge University Press, Cambridge, UK, 2004).
- [17] V. Batagelj and A. Mrvar, (2006), Data available at <http://vlado.fmf.uni-lj.si/pub/networks/data/default.htm>.
- [18] M. E. J. Newman, *Physical Review E* **74**, 036104 (2006).
- [19] R. R. Faulkner, *Music on Demand. Composers and Careers in the Hollywood Film Industry* (Transaction Books, New Brunswick, NJ, USA, 1983).
- [20] P. Gleiser and L. Danon, *Advances in Complex Systems* **6**, 565 (2003).
- [21] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *SIAM Review* **53**, 526 (2011).
- [22] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Physical Review E* **68**, 056110 (2003).
- [23] D. J. Fenn, (2008), Data downloaded on 19 December 2008 from <http://uk.finance.yahoo.com/>.
- [24] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, *Science* **327**, 439 (2010).
- [25] M. D. Fricker, L. Boddy, and D. P. Bebber, *Topologica* **1**, 4 (2008).
- [26] L. L. M. Heaton, E. López, P. K. Maini, M. D. Fricker, and N. S. Jones, *Proceedings of the Royal Society B* **277**, 3265 (2010).
- [27] D. P. Bebber, J. Hynes, P. R. Darrah, L. Boddy, and M. D. Fricker, *Proceedings of the Royal Society B* **274**, 2307 (2007).
- [28] V. Batagelj, A. Mrvar, and M. Zaveršnik, in *Proceedings of the 5th International Multi-Conference Information Society – Language Technologies*, edited by T. Erjavec and J. Gros (2002) pp. 143–148.
- [29] J. M. Reitz, (2000), ODLIS is maintained by Joan Reitz and was converted into the network used here by A. Mrvar and V. Batagelj. The data is available at <http://vlado.fmf.uni-lj.si/pub/networks/data/dic/odlis/Odlis.htm>.
- [30] J. C. Johnson and L. Krempel, *Journal of Social Structures* **5** (2004).
- [31] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing* (Addison-Wesley, Reading, MA, USA, 1993).
- [32] R. Milo, S. Itzkovitz, N. Kashtan, R. Levittand, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [33] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, *Nature* **407**, 651 (2000).
- [34] V. Krebs, (2004), Network compiled by Valdis Krebs and not previously published. The data is available at <http://www.orgnet.com/>.
- [35] V. Batagelj, (2003), Data available at <http://vlado.fmf.uni-lj.si/pub/networks/data/2mode/journals.htm>.
- [36] S. Mangan and U. Alon, *Proceedings of the National Academy of Sciences of the U.S.A.* **100**, 11980 (2003).
- [37] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [38] T. Callaghan, P. J. Mucha, and M. A. Porter, *American Mathematical Monthly* **114**, 761 (2007).
- [39] M. E. J. Newman, (2006), Network compiled by Mark Newman and not previously published. The data is available at <http://www-personal.umich.edu/~mejn/netdata/>.
- [40] E. Garfield, I. H. Sher, and R. J. Torpie, *The Use of Citation Data in Writing the History of Science* (The Institute for Scientific Information, Philadelphia, PA, USA, 1964).
- [41] S. Gopal, “Societies and Cities in the Age of Instant Access,” (Springer, Berlin, Germany, 2007).
- [42] L. A. Adamic and N. Glance, in *LinkKDD ’05: Proceedings of the 3rd international workshop on Link discovery* (2005) pp. 36–43.
- [43] D. F. Gleich, (2001), Data available at <http://www.cise.ufl.edu/research/sparse/matrices/Gleich/wb-cs-stanford.html>.
- [44] C. A. Hidalgo, B. Klinger, A. L. Barabási, and R. Hausmann, *Science* **317**, 482 (2007).
- [45] D. A. Smith and D. R. White, *Social Forces* **70**, 857 (1992).
- [46] J. H. Fowler, *Social Networks* **28**, 454 (2006).
- [47] J. H. Fowler, *Political Analysis* **14**, 456 (2006).
- [48] M. A. Porter, P. J. Mucha, M. E. J. Newman, and C. M. Warmbrand, *Proceedings of the National Academy of Sciences of the U.S.A.* **102**, 7057 (2005).
- [49] M. A. Porter, P. J. Mucha, M. E. J. Newman, and A. J. Friend, *Physica A* **386**, 414 (2007).
- [50] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter, arXiv:0907.3509 (2010).

- [51] K. T. Poole and H. Rosenthal, *Congress: A Political-Economic History of Roll Call Voting* (Oxford University Press, Oxford, UK, 1997).
- [52] N. M. McCarty, K. T. Poole, and H. Rosenthal, *Polarized America: The Dance of Ideology and Unequal Riches* (MIT Press, Cambridge, MA, USA, 2007).
- [53] D. Firth and A. Spirling, Working paper (2003), available at <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic/firth/software/tapir/firth-spirling.pdf>.
- [54] E. Voeten, *International Organization* **54**, 185 (2000).
- [55] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, *Nucleic Acids Research* **34**, D535 (2006).
- [56] L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, and D. Eisenberg, *Nucleic Acids Research* **32**, D449 (2004).
- [57] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg, *Nucleic acids research* **30**, 303 (2002).
- [58] J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, and et al., *Nature* **437**, 1173 (2005).
- [59] K. Brown and I. Jurisica, *Genome Biology* **8**, R95 (2007).
- [60] K. R. Brown and I. Jurisica, *Bioinformatics* **21**, 2076 (2005).
- [61] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, and C. von Mering, *Nucleic Acids Research* **37**, D412 (2009).
- [62] S. Agarwal, C. M. Deane, M. A. Porter, and N. S. Jones, *PLoS Computational Biology* **6**, e1000817 (2010).
- [63] N. Bertin, N. Simonis, D. Dupuy, M. E. Cusick, J. D. J. Han, H. B. Fraser, F. P. Roth, and M. Vidal, *PLoS Biology* **5**, e153 (2007).
- [64] J. D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, *Nature* **430**, 88 (2004).
- [65] K. Tarassov, V. Messier, C. R. Landry, S. Radinovic, M. M. S. Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey, and S. W. Michnick, *Science* **320**, 1465 (2008).
- [66] J. Scott and M. Hughes, *The anatomy of Scottish capital: Scottish companies and Scottish capital, 1900-1979* (Croom Helm, London, UK, 1980).
- [67] K. Norlen, G. Lucas, M. Gebbie, and J. Chuang, in *Proceedings of International Telecommunications Society 14th Biennial Conference (ITS2002)* (2002).
- [68] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, *Behavioral Ecology and Sociobiology* **54**, 396 (2003).
- [69] E. M. Rogers and D. L. Kincaid, *Communication Networks. Toward a New Paradigm for Research* (The Free Press, New York, NY, USA, 1981).
- [70] D. Krackhardt, *Research in the Sociology of Organizations* **16**, 183 (1999).
- [71] J. H. Michael and J. G. Massey, *Forest Products Journal* **47**, 25 (1997).
- [72] R. Alberich, J. Miro-Julia, and F. Rossello, arXiv:cond-mat/0202174 (2002).
- [73] J. Gil-Mendieta and S. Schmidt, *Social Networks* **18**, 355 (1996).
- [74] M. Boguná, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, *Physical Review E* **70**, 056122 (2004).
- [75] H. R. Bernard, P. D. Killworth, and L. Sailer, *Social Networks* **2**, 191 (1979-1980).
- [76] H. R. Bernard, P. D. Killworth, and L. Sailer, *Social Science Research* **11**, 30 (1982).
- [77] A. K. Romney and S. C. Weller, *Social Networks* **6**, 59 (1984).
- [78] H. R. Bernard and P. D. Killworth, *Human Organization* **35**, 269 (1976).
- [79] H. R. Bernard and P. D. Killworth, *Human Communication Research* **4**, 3 (1977).
- [80] P. D. Killworth and H. R. Bernard, *Social Networks* **2**, 19 (1979-1980).
- [81] B. Kapferer, *Strategy and transaction in an African factory* (Manchester University Press, Manchester, UK, 1972).
- [82] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, *Physical Review E* **68**, 065103 (2003).
- [83] W. W. Zachary, *Journal of Anthropological Research* **33**, 452 (1977).
- [84] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, *Physical Review Letters* **96**, 114102 (2006).